MẠNG LƯỚI QUY ĐỊNH DI TRUYỀN VỀ KÍCH THƯỚC HẠT ĐẬU TƯƠNG, HÀM LƯỢNG DẦU VÀ PROTEIN

Zongbiao Duan[1,2], Qing Li[3], Hong Wang[3], Xuemei He[2] và Min Zhang[2]
Võ Như Cầm biên dịch.

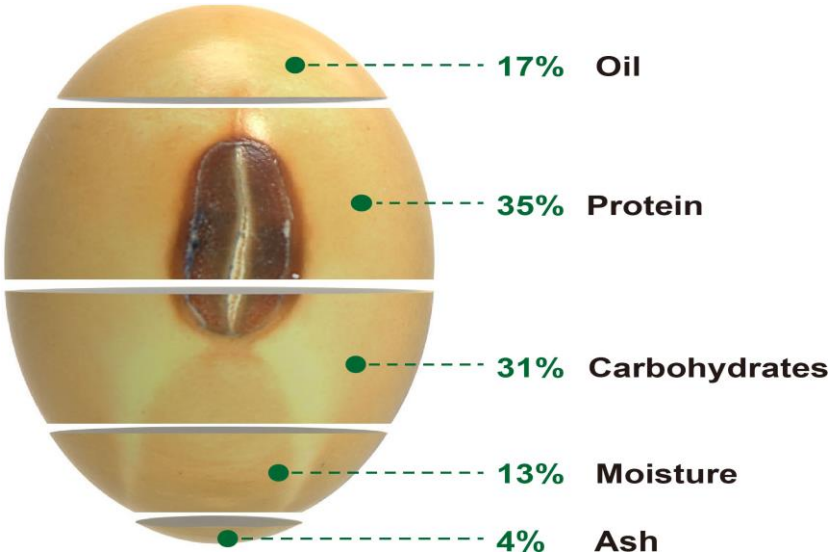
1. Phòng thí nghiệm hạt giống Vịnh Yazhou Hải Nam, Tam Á, Trung Quốc.
2. Phòng thí nghiệm Trọng điểm quốc gia về Tế bào thực vật và Kỹ thuật Nhiễm sắc thể, Viện Di truyền và Sinh học Phát triển, Viện Hàn lâm Khoa học Trung Quốc, Bắc Kinh, Trung Quốc.
3. Phòng Thí nghiệm Trọng điểm quốc gia về Sinh học và Chọn giống lúa, Viện Nghiên cứu lúa gạo quốc gia Trung Quốc, Viện Khoa học Nông nghiệp Trung Quốc, Hàng Châu, Trung Quốc.

Là cây lấy hạt có dầu hàng đầu cung cấp dầu thực vật và protein cho đời sống hàng ngày của con người, tăng năng suất và nâng cao chất lượng dinh dưỡng (nhiều dầu hoặc nhiều đạm) là hai mục tiêu cơ bản hàng đầu của chọn tạo giống đậu tương. Kích thước hạt là một trong những yếu tố quan trọng nhất quyết định năng suất đậu tương. Kích thước hạt, hàm lượng dầu và protein là những tính trạng số lượng phức tạp bị chi phối bởi các yếu tố di truyền và môi trường trong quá trình phát triển hạt. Thành phần và số lượng chất dự trữ hạt ảnh hưởng trực tiếp đến kích thước hạt. Nhìn chung, dầu và protein chiếm gần 60% tổng lượng hạt đậu tương dự trữ. Do đó, kích thước hạt, dầu hoặc hàm lượng protein của đậu tương là những đặc điểm nông học có mối tương quan cao. Tăng kích thước hạt giúp tăng năng suất đậu tương và có thể cải thiện chất lượng hạt giống. Tương tự như vậy, hàm lượng dầu và protein tăng lên sẽ cải thiện chất lượng dinh dưỡng của đậu tương và có thể sẽ làm tăng năng suất đậu tương. Do tầm quan trọng của ba tính trạng hạt giống này trong chọn giống đậu tương, các nghiên cứu sâu rộng đã được tiến hành trên locus tính trạng số lượng (QTL) hoặc gen cơ bản của chúng và phân tích các con đường điều hòa phân tử của chúng. Đánh giá này đã tóm tắt tiến trình trong bộ gen chức năng kiểm soát kích thước hạt đậu tương, hàm lượng dầu và protein trong những thập kỷ gần đây, đồng thời đưa ra những thách thức và triển vọng để phát triển các giống đậu tương năng suất cao với hàm lượng dầu hoặc protein cao. Cuối cùng, chúng tôi hy vọng bài đánh giá này sẽ giúp ích cho việc cải thiện năng suất và chất lượng đậu tương trong quá trình nhân giống sau này.

1. Giới thiệu

Dầu và protein là những chất dinh dưỡng thiết yếu cho con người và gia súc, với gần 70% dầu ăn và một nửa protein thức ăn chăn nuôi đến từ thực vật. Đậu tương (*Glycine max*) cung cấp gần 60% sản lượng hạt có dầu toàn cầu và chiếm hơn 25% lượng protein tiêu thụ cho thực phẩm và thức ăn chăn nuôi trên toàn thế giới, khiến nó trở thành cây trồng thương mại hàng đầu để sản xuất dầu thực vật và protein (Wang và cs, 2020b). Đậu tương trồng trọt được thuần hóa từ đậu tương dại (*Glycine soja*) ở miền trung Trung Quốc khoảng 5.000 năm trước và sau đó lan rộng ra khắp thế giới (Carter và cs, 2004; Wilson, 2008). Là cây lấy hạt có dầu và thức ăn gia súc chiếm ưu thế, hạt đậu tương được canh tác hiện đại chứa khoảng 17% dầu, 35% protein (bao gồm cả axit amin thiết yếu và không thiết yếu), 31% carbohydrate (bao gồm cả carbohydrate hòa tan và không hòa tan), 13% độ ẩm và 4% tro (Liu, 1997) (Hình 1). Hàm lượng dầu của hạt đậu tương dao động từ 8,3–27,9% và hàm lượng protein thay đổi từ 34,1–56,8% tùy thuộc vào giống đậu tương và điều kiện canh tác (Wilson, 2004). Dầu đậu tương được tạo ra và lưu trữ chủ yếu dưới dạng axit béo (FA), triacylglycerol (TAG) và tocopherol (Liu và cs, 2022). Có năm FA trung tâm trong hạt đậu tương, bao gồm axit stearic (C18:0), axit oleic (C18:1),

axit linoleic (C18:2), axit linolenic (C18:3) và axit palmitic (C16:0), có thành phần quyết định trực tiếp đến chất lượng dầu đậu tương. Protein hạt đậu tương bao gồm chủ yếu là các protein dự trữ như glycinin (11S globulin) và conglycinin (7S globulin) (Liu và cs, 2022).



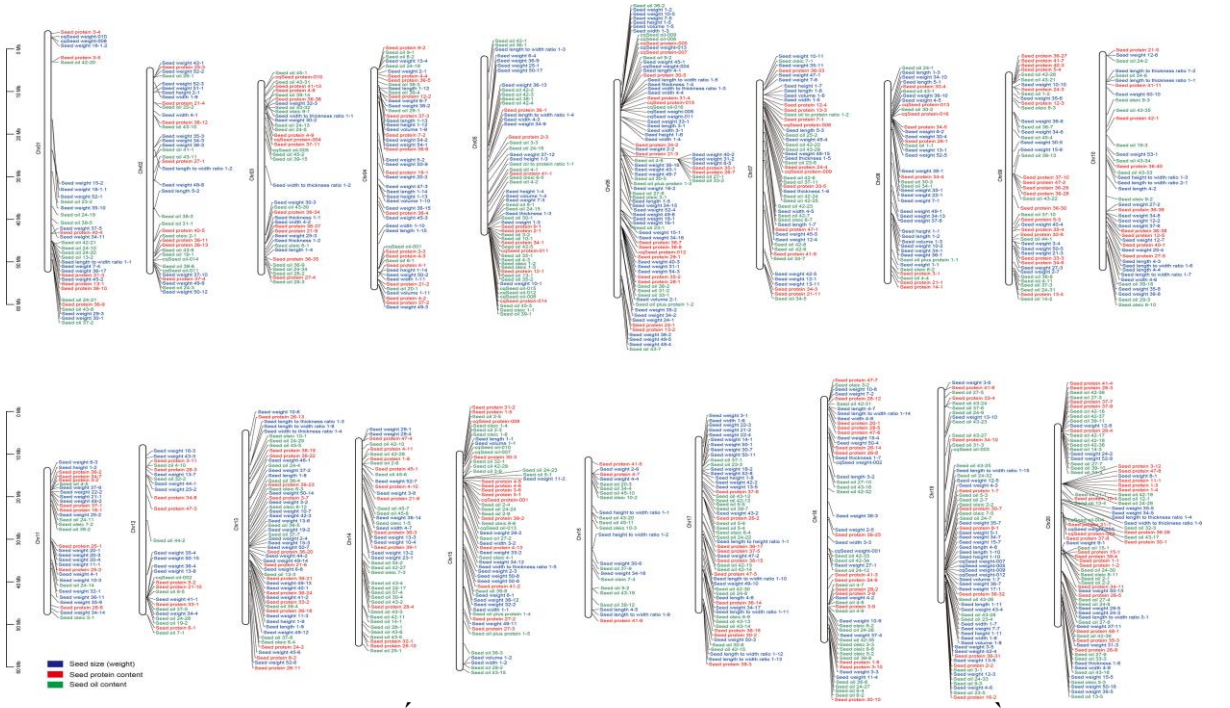
Hình 1. Thành phần của hạt đậu tương trưởng thành được bảo quản. Giá trị phần trăm cho biết trọng lượng tương đối của thành phần tương ứng trong hạt (Liu, 1997).

Những tiến bộ gần đây đã chỉ ra rằng năng suất cây trồng toàn cầu cần phải tăng gấp đôi vào năm 2050 để đáp ứng nhu cầu tiêu thụ và dân số ngày càng tăng (Godfray và cs, 2010; Tilman và cs, 2011), có nghĩa là sản lượng cây trồng tăng 2,4% mỗi năm. Tuy nhiên, sản xuất đậu tương tụt hậu nghiêm trọng so với nhu cầu dự kiến, chỉ tăng trung bình 1,3% mỗi năm (Ray và cs, 2013). So với các loại cây lương thực, bao gồm gạo, lúa mì và ngô, năng suất đậu tương chỉ bằng khoảng một phần ba đến một nửa. Vì vậy, nâng cao năng suất đậu tương là nhiệm vụ cần thiết và cấp bách đối với công tác chọn tạo giống đậu tương. Tăng kích thước hạt là một trong những cách quan trọng để tăng năng suất đậu tương. Kích thước hạt đậu tương có thể được mô tả bằng chiều dài (đường kính song song với rốn), chiều rộng (đường kính từ rốn đến bề mặt trục của hạt), và độ dày (đường kính dọc theo rốn), và thành phần và hàm lượng dự trữ hạt giống trực tiếp xác định nó. Đậu tương trồng trọt thường tạo ra hạt to hơn với mức dầu cao hơn (Wang và cs, 2020b). Đậu tương dại có hạt nhỏ hơn với hàm lượng dầu thấp hơn đậu tương trồng. Tuy nhiên, hàm lượng protein trong hạt không tăng ở các giống đậu tương hạt lớn (Wang và cs, 2020b). Do đó, cải thiện đậu tương liên quan đến việc tăng song song kích thước hạt, tích lũy dầu và có thể kèm theo sự thay đổi về mức độ protein.

Trong nhiều thập kỷ, việc tăng kích thước hạt, tích lũy dầu và hàm lượng protein là mục tiêu thiết yếu của các chương trình chọn giống đậu tương. Việc công bố bộ gen tham chiếu của đậu tương (Williams 82) vào năm 2010 đã thúc đẩy mạnh mẽ sự phát triển của bộ gen chức năng đậu tương (Zhang và cs, 2022). Ở đây, chúng tôi xem xét những tiến bộ trong bộ gen chức năng đậu tương về kích thước hạt, tích lũy dầu và hàm lượng protein. Ngoài ra, chúng tôi cũng thảo luận về những thách thức và triển vọng phát triển các giống đậu tương năng suất cao với hàm lượng dầu hoặc protein cao. Vì sự tổng hợp sinh hóa của dầu trong hạt đã được nghiên cứu rộng rãi và đánh giá tốt (Bates và cs, 2013; Xu và Shanklin, 2016; Song và cs, 2017; Liu và cs, 2022; Yang và cs, 2022a), chúng tôi sẽ không nhắc lại những bình luận này ở đây.

2. Lập bản đồ gen liên quan đến kích thước hạt, hàm lượng dầu và protein

Kích thước hạt, hàm lượng dầu và protein là những đặc điểm phức tạp được kiểm soát bởi các yếu tố di truyền và môi trường trong quá trình phát triển và trưởng thành của hạt. Do tầm quan trọng của chúng trong chọn tạo giống đậu tương, các nhà nghiên cứu đã thực hiện phân tích liên kết mở rộng để xác định các locus tính trạng số lượng (QTL) liên quan đến ba tính trạng hạt này bằng cách sử dụng các quần thể có nguồn gốc từ bố mẹ khác nhau, chẳng hạn như quần thể F_2, dòng lai tái tổ hợp (RIL), thay thế đoạn nhiễm sắc thể dòng (CSSL) và dòng gần đồng phân (NIL) (Han và cs, 2012; Eskandari và cs, 2013a; Eskandari và cs, 2013b; Qi và cs, 2014; Warrington và cs, 2015; Wang và cs, 2015a; Yang và cs, 2019; Cui và cs, 2020; Kumawat và Xu, 2021; Kumar và cs, 2022; Luo và cs, 2022; Yang và cs, 2022b). Cho đến nay, hàng trăm QTL liên quan đến kích thước hạt (bao gồm trọng lượng hạt), tích lũy dầu và hàm lượng protein đã được ghi lại trong Cơ sở dữ liệu bộ gen của SoyBase (//www.soybase.org). Chẳng hạn, có 396 QTL cho kích thước và trọng lượng hạt (Hình 2; Bảng bổ sung 1), 333 QTL cho hàm lượng dầu trong hạt (Hình 2; Bảng bổ sung 2) và 234 QTL cho hàm lượng protein trong hạt (Hình 2; Bảng bổ sung 3). Trong số các QTL này, một số QTL liên quan đến kích thước hạt, sự tích lũy dầu và hàm lượng protein chia sẻ các vùng chồng lấn, cho thấy sự hiện diện của các gen điều hòa đa hướng trong các QTL này. Tuy nhiên, do các chỉ thị phân tử có độ phân giải thấp và mật độ thấp cũng như kích thước quần thể hạn chế, hầu hết các QTL được lập bản đồ trong một vùng nhiễm sắc thể lớn, làm cho các QTL này kém hiệu quả hơn trong việc xác định gen cụ thể để cải thiện cây trồng. Hiện tại, chỉ có một số gen liên quan đến kích thước hạt, tích lũy dầu và hàm lượng protein đã được phân lập từ ánh xạ QTL, chẳng hạn như *GmPP2C-1* (Lu và cs, 2017), *GmB1* (Zhang và cs, 2018a) và *Glyma.20G85100* (còn được gọi là *GmSWEET39*) (Zhang và cs, 2020; Fliege và cs, 2022). Ngoài ra, hai gen liên quan đến kích thước/trọng lượng hạt đã được xác định thông qua phân tích nhân bản gen dựa trên bản đồ đột biến hoặc phép lai so sánh bộ gen (CGH), bao gồm *GmSSS1* (Zhu và cs, 2022) và *GmKIX8-1* (Nguyen và cs, 2021).



Hình 2. Các QTL liên quan đến kích thước hạt (trọng lượng), tích lũy dầu và hàm lượng protein trong đậu tương. Các QTL này được lấy từ cơ sở dữ liệu SoyBase (https://soybase.org/).

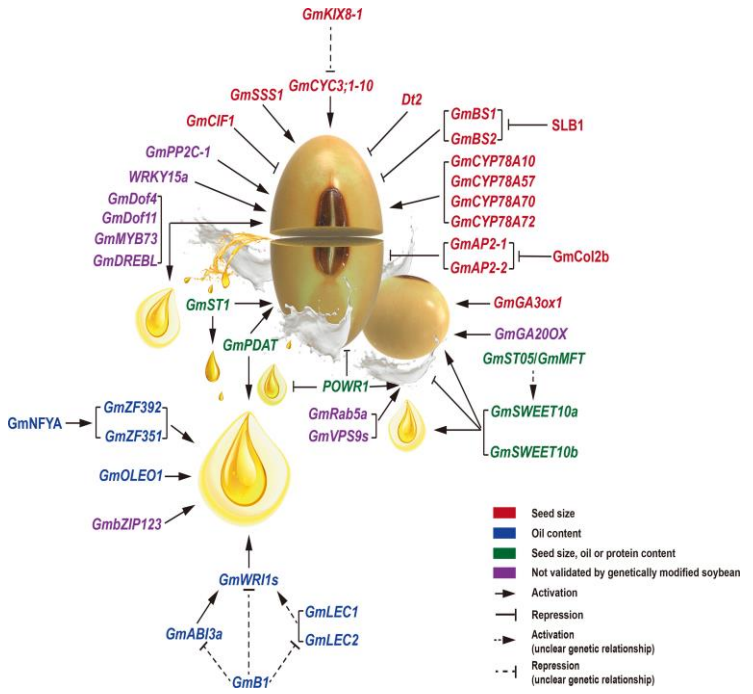
Với sự phát triển của omics, nghiên cứu kết hợp trên toàn bộ bộ gen (GWAS) đã trở thành một phương pháp lập bản đồ gen hoặc QTL mạnh mẽ để phân tích các đặc điểm nông học phức tạp ở cây trồng. So với phân tích liên kết hoặc lập bản đồ QTL thông thường, GWAS mang lại những lợi thế đáng kể: 1) GWAS không cần xây dựng quần thể lập bản đồ. 2) Quần thể GWAS bao gồm nhiều biến thể tự nhiên hơn so với quần thể có hai bố mẹ. 3) GWAS có thể đạt được độ phân giải bản đồ cao hơn nhờ các chỉ thị phân tử mật độ cao và các sự kiện tái tổ hợp lịch sử đa dạng (Wang và cs, 2020a; Li và cs, 2022b). Trong thập kỷ qua, hàng chục GWAS đã được thực hiện để xác định các QTL hoặc nucleotide tính trạng định lượng (QTN) liên quan đến kích thước hạt, tích lũy dầu và mức protein trong đậu tương (Hwang và cs, 2014; Zhou và cs, 2015; Zhang và cs, 2016a; Yan và cs, 2017; Zhang và cs, 2018b; Lee và cs, 2019; Zhao và cs, 2019; He và cs, 2021; Zhang và cs, 2021; Hong và cs, 2022). Dựa trên phương pháp này, *GmOLEO1* (Zhang và cs, 2019b), *GmPDAT* (Liu và cs, 2020a), *GmSWEET10a* (còn được gọi là *GmSWEET39*) (Miao và cs, 2020; Wang và cs, 2020b) và *GmST05* (Duan và cs, 2022) đã được xác định và xác nhận có liên quan đến những đặc điểm hạt giống này, cho thấy cách này hiệu quả hơn. Mặc dù GWAS có lợi thế trong việc lập bản đồ di truyền, nhưng cấu trúc quần thể và các mối quan hệ cá nhân có khả năng tạo ra kết quả dương tính giả trong phân tích liên kết. Do đó, tốt hơn là tích hợp bản đồ liên kết và phân tích GWAS để phân tích các đặc điểm phức tạp. Các phương pháp phân tích hỗn hợp đã sử dụng và lập bản đồ thành công các QTL hoặc QTN liên quan đến các đặc điểm hạt này ở đậu tương (Cao và cs, 2017; Zhang và cs, 2019c), đồng thời nhân bản thêm *GmSWEET39* (Zhang và cs, 2020), *GmGA3ox1* (Hu và cs, 2022), *GmST1* (Li và cs, 2022a) và *POWR1* (Goettel và cs, 2022).

3. Gen quy định kích thước hạt

Hạt của thực vật bậc cao bao gồm phôi, nội nhũ và vỏ hạt, trong đó phôi và nội nhũ lần lượt được tạo ra từ tế bào trứng đã thụ tinh và tế bào trung tâm. Ngược lại, vỏ hạt được phát triển từ vỏ bào tử. Do đó, kích thước hạt được xác định bởi các tín hiệu tích hợp của mô mẹ và mô hợp tử ảnh hưởng đến sự phát triển phối hợp của phôi, nội nhũ và vỏ hạt (Li và cs, 2019). Một số con đường truyền tín hiệu kiểm soát kích thước hạt của cây mẹ đã được xác định ở cây Arabidopsis và cây lúa, chẳng hạn như tín hiệu G-protein, tín hiệu ubiquitin-proteasome, tín hiệu protein kinase hoạt hóa bằng mitogen (MAPK), tín hiệu phytohormone và một số chất điều hòa phiên mã. Trong khi đó, con đường HAIKU (IKU) và một số kích thích tố thực vật điều chỉnh một phần sự phát triển của các mô hợp tử (Li và cs, 2019). Tuy nhiên, so với cây Arabidopsis và cây lúa, mạng lưới phân tử điều chỉnh kích thước hạt ở cây đậu tương vẫn còn kém.

Là thành phần điều hòa quan trọng của biểu hiện gen, một số yếu tố phiên mã (TF) liên quan đến kích thước hạt đã được xác định ở đậu tương (Hình 3; Bảng 1). BIG SEEDS1 (BS1) thuộc thành viên nhóm II của họ lớn TIFY TF. Nó đóng một vai trò quan trọng trong việc kiểm soát kích thước của hạt, quả và lá thông qua một mô-đun điều tiết nhằm mục tiêu tăng sinh tế bào trong cây họ đậu mô hình của *Medicago truncatula* (Ge và cs, 2016). Điều chỉnh giảm các gen chỉnh hình BS1 (*GmBS1* và *GmBS2*) trong đậu tương dẫn đến tăng kích thước hạt và hàm lượng axit amin. SLB1 mã hóa protein hộp F tạo thành một phần của phức hợp ligase ubiquitin SKP1/Cullin/F-box E3. Các phân tích sinh hóa và di truyền cho thấy SLB1 tương tác với BS1 để kiểm soát sự phân nhánh bên và sự phát triển của cơ quan bằng cách điều chỉnh sự ổn định của protein BS1 trong *Medicago truncatula*. Ngoài ra, sự biểu hiện quá mức của SLB1 dẫn đến kích thước lá và hạt tăng

lên ở cả *Medicago truncatula* và đậu tương, cho thấy sự bảo tồn chức năng của SLB1 (Yin và cs, 2020). TF WRKY thực vật tham gia vào nhiều quá trình sinh học, chẳng hạn như quá trình tạo phôi và phát triển hạt giống (Luo và cs, 2005). *WRKY15a* được thể hiện một cách khác biệt trong quá trình phát triển quả giữa đậu tương trồng và đậu tương dại. Bốn haplotypes (H1-H4) đã có mặt trong *WRKY15a*, khác nhau trong quỹ tích kính hiển vi lõi CT tại vùng 5' chưa được dịch mã (5'-UTR) của *WRKY15a*. Kiểu haplotype H1 với sáu lần lặp lại CT là alen duy nhất trong đậu tương trồng, trong khi kiểu haplotype H3 với năm lần lặp lại CT là alen chính trong đậu tương dại. Trọng lượng hạt với haplotype H1 nặng hơn so với đậu tương dại chứa haplotypes H2, H3 và H4, và trọng lượng hạt có mối tương quan thuận với biểu hiện *WRKY15a*, cho thấy tác động tích cực của *WRKY15a* đối với kích thước hạt (Gu và cs, 2017). *Dt2*, mã hóa TF hộp MADS, đóng vai trò thiết yếu trong việc kiểm soát nhiều đặc điểm nông học, chẳng hạn như thời gian ra hoa, thói quen phát triển của thân và chiều cao cây (Ping và cs, 2014; Zhang và cs, 2019a). Một báo cáo gần đây đã chỉ ra rằng *Dt2* cũng xác định sự phân nhánh của chồi và kích thước hạt (Liang và cs, 2022). Các dòng loại bỏ *Dt2* thực hiện nhiều thay đổi tính trạng liên quan đến năng suất, chẳng hạn như chiều dài và chiều rộng hạt tăng lên, trọng lượng hạt nặng hơn và trọng lượng hạt trên mỗi cây cao hơn, do đó dẫn đến năng suất trên mỗi ô được cải thiện rõ rệt. Ngược lại, các dòng biểu hiện quá mức *Dt2* biểu hiện chiều dài và chiều rộng hạt giảm.



Hình 3. Mạng lưới di truyền quy định kích thước hạt (trọng lượng), tích lũy dầu và hàm lượng protein trong đậu tương. Các gen hoặc protein liên quan đến kích thước hạt (trọng lượng) và hàm lượng dầu được hiển thị tương ứng bằng phông chữ màu đỏ và màu xanh lam. Các bộ điều chỉnh đa hướng cho kích thước hạt (trọng lượng), tích lũy dầu hoặc hàm lượng protein được biểu thị bằng phông chữ màu xanh lá cây. Các gen điều hòa, có chức năng chỉ được xác nhận ở cây Arabidopsis chứ không phải cây đậu tương, được hiển thị bằng phông chữ màu tím.

Một số gen mã hóa các enzym khác nhau cũng đã được chứng minh là có ảnh hưởng đến kích thước hạt đậu tương (Hình 3; Bảng 1). Gen phosphatase 2C-1 (*GmPP2C-1*) từ đậu tương dại giúp tăng trọng lượng hoặc kích thước hạt bằng cách cải thiện kích thước tế bào vỏ và kích hoạt một tập hợp con các gen liên quan đến tính trạng của hạt (Lu và cs, 2017). Ngoài ra, *GmPP2C-1* tạo điều kiện tích lũy protein *GmBZR1* khử phospho, đóng vai trò là yếu tố phiên mã chính trong tín hiệu BR. Hơn nữa, sự biểu hiện quá mức của *GmBZR1* có thể cải thiện kích thước hoặc trọng lượng hạt ở cây Arabidopsis chuyển gen. Invertase của thành tế bào (CWI) đóng một vai trò quan trọng trong việc truyền tín hiệu và chuyển hóa đường, ảnh hưởng đến sự tương tác nguồn chìm và sự phát triển của hạt

(Tang và cs, 2017). *GmCIF1* mã hóa chất ức chế invertase của thành tế bào và việc ức chế biểu hiện gen *GmCIF1* cho thấy các hoạt động CWI gia tăng và kích thước hạt lớn hơn trong khi tích lũy nhiều protein, hexose và tinh bột hơn trong hạt đậu tương.

Bảng 1. Các gen đại diện liên quan đến kích thước hạt, tích lũy dầu và hàm lượng protein trong đậu tương.

Gene	Accession No. (Wm82.a2.v1)	Functional annotation	Controlling trait	References
GmCYP78A10	Glyma.05G019200	cytochrome P450 CYP2 subfamily	seed size/weight	Wang et al., 2015b
GmCYP78A57	Glyma.02G119600	cytochrome P450 CYP2 subfamily/homolog of AtCYP78A6/KLU	seed size/weight	Zhao et al., 2016
GmCYP78A70	Glyma.01G061100	cytochrome P450 CYP2 subfamily/homolog of AtCYP78A7/KLU	seed size/weight	Zhao et al., 2016
GmCYP78A72	Glyma.19G240800	cytochrome P450 CYP2 subfamily/homolog of AtCYP78A5/KLU	seed size/weight	Zhao et al., 2016
GA20OX	Glyma.07G081700	gibberellin 20 oxidase	seed size/weight	Lu et al., 2016
GmBS1	Glyma.10G244400	a group II member of the TIFY family of transcription regulators	seed size/weight	Ge et al., 2016
GmBS2	Glyma.20G150000	a group II member of the TIFY family of transcription regulators	seed size/weight	Ge et al., 2016
WRKY15a	Glyma.05G096500	WRKY transcription factor	seed size/weight	Gu et al., 2017
GmPP2C-1	Glyma.17G221100	a putative phosphatase 2C protein	seed size/weight	Lu et al., 2017
GmCIF1	Glyma.17G036300	cell-wall inhibitor of beta-fructosidase	seed size/weight	Tang et al., 2017
GmKIX8-1	Glyma.17G112800	unknow protein	seed size/weight	Nguyen et al., 2021
GmSSS1	Glyma.19G196000	SPINDLY homolog protein	seed size/weight	Zhu et al., 2022
GmGA3ox1	Glyma.07G033800	gibberellin 3β-hydroxylase	seed size/weight	Hu et al., 2022
Dt2	Glyma.18G273600	MADS-box transcription factor	seed size/weight	Liang et al., 2022
GmCOL2b	Glyma.19G039000	B-box type zinc finger protein with CCT domain	seed size/weight	Yu et al., 2023
GmAP2-1	Glyma.01g188400	AP2 transcription factor	seed size/weight	Yu et al., 2023
GmAP2-2	Glyma.05g091200	AP2 transcription factor	seed size/weight	Yu et al., 2023
GmbZIP123	Glyma.06G010200	bZIP transcription factor	oil content	Song et al., 2013
GmNFYA	Glyma.02G303800	a domain of the nuclear transcription factor Y	oil content	Lu et al., 2016
GmLEC1	Glyma.07G268100	nuclear transcription factor Y subunit B	oil content	Pelletier et al., 2017
GmLEC2a	Glyma.20G035800	AP2/B3-like transcriptional factor	oil content	Manan et al., 2017
GmLEC2b	Glyma.20G035700	AP2/B3-like transcriptional factor	oil content	Manan et al., 2017
GmZF351	Glyma.06G290100	zinc finger C-x8-C-x5-C-x3-H type family protein	oil content	Li et al., 2017
GmWRI1a	Glyma.15G221600	ethylene-responsive transcription factor WRI1a	oil content	Chen et al., 2018
GmB1	Glyma.13G241700	a transmembrane transporter-like protein	oil content	Zhang et al., 2018a
GmOLEO1	Glyma.20G196600	oleosin-encoding gene	oil content	Zhang et al., 2019b
GmPDAT	Glyma.13G108100	regulator of chromosome condensation	oil content	Liu et al., 2020a
GmWRI1b	Glyma.08G227700	ethylene-responsive transcription factor WRI1b	oil content	Guo et al., 2020
GmZF392	Glyma.12G205700	zinc finger C-x8-C-x5-C-x3-H type family protein	oil content	Lu et al., 2021
Rab5a2	Glyma.13G153000	a small GTPase	protein content	Wei et al., 2020
	Glyma.18G045000	a small GTPase	protein content	
GmDof4	Glyma.17G081800	Dof zinc finger protein	seed size/weight and oil content	Wang et al., 2007
GmDof11	Glyma.13G329000	Dof zinc finger protein	seed size/weight and oil content	Wang et al., 2007
GmMYB73	Glyma.06G303100	MYB transcription factor MYB73	seed size/weight and oil content	Liu et al., 2014
GmDREBL	Glyma.12G103100	DREB-type transcription factors	seed size/weight and oil content	Zhang et al., 2016b
GmST1	Glyma.08G109100	UDP-D-glucuronate 4-epimerase	seed size/weight and oil content	Li et al., 2022a
GmSWEET10a	Glyma.15G049200	sugar efflux transporter SWEET39	seed size/weight, oil and protein contents	Wang et al., 2020b; Miao et al., 2020
GmSWEET10b	Glyma.08G183500	sugar efflux transporter SWEET24	seed size/weight, oil and protein contents	Wang et al., 2020b
GmST05	Glyma.05G244100	phosphatidylethanolamine-binding protein	seed size/weight, oil and protein contents	Duan et al., 2022
POWR1	Glyma.20G085100	CCT domain protein	seed size/weight, oil and protein contents	Fliege et al., 2022; Goettel et al., 2022

GmSSS1 mã hóa một O-GlcNAc transferase giả định trong đậu tương. Loại bỏ *GmSSS1* dẫn đến hạt nhỏ, trong khi *GmSSS1* biểu hiện quá mức tạo ra hạt lớn (Zhu và cs, 2022). Điều chỉnh *GmSSS1* có thể ảnh hưởng tích cực đến sự phân chia và mở rộng tế bào ở cây chuyển gen. *GmGA3ox1*, một gibberellin (GA) 3β-hydroxylase trong đậu tương, là enzyme quan trọng trong con đường sinh tổng hợp GA. Loại bỏ *GmGA3ox1* dẫn đến giảm sinh tổng hợp GA trong khi tăng cường quang hợp (Hu và cs, 2022). Cây loại bỏ *GmGA3ox1* cho thấy trọng lượng và chiều dài hạt giảm, nhưng sản lượng hạt được cải thiện bằng cách tăng số nhánh, quả và hạt. Ngược lại, sự biểu hiện quá mức của *GmGA3ox1* làm tăng trọng lượng và chiều dài hạt ở đậu tương chuyển gen. Tương tự, sự biểu hiện quá mức của *GA20OX*, mã hóa gibberellin 20 oxidase trong bước giới hạn tốc độ của quá trình sinh tổng hợp GA, đã nâng cao kích thước/trọng lượng hạt của cây Arabidopsis chuyển gen (Lu và cs, 2016).

Bên cạnh các gen trên, một số gen đậu tương tương đồng được biết là điều chỉnh kích thước hạt ở cây Arabidopsis cũng đã được chứng minh là kiểm soát kích thước hạt đậu tương (Hình 3; Bảng 1). Ví dụ, một số thành viên họ *P450/CYP78A* được đề xuất để kiểm soát kích thước hạt ở Arabidopsis (Wang và cs, 2008; Fang và cs, 2012). Các ortholog *P450/CYP78A* trong đậu tương, chẳng hạn như *GmCYP78A10*, *GmCYP78A57*, *GmCYP78A70* và *GmCYP78A72*, thể hiện chức năng được bảo tồn để cải thiện kích thước hoặc trọng lượng hạt (Wang và cs, 2015b; Zhao và cs, 2016; Du và cs, 2017) , nhưng cơ chế cơ bản về cách thức hoạt động của chúng phần lớn vẫn còn khó nắm bắt. Phức hợp ức chế PPD/KIX/TPL bao gồm các protein PPD2, KIX8/9 và TPL đã được chứng minh là có ảnh hưởng đến kích thước cơ quan bằng cách điều chỉnh sự tăng sinh mô phân sinh ở cây *Arabidopsis* (Baekelandt và cs, 2018). *GmKIX8-1*, một ortholog *AtKIX8* đậu tương, cũng tham gia vào việc kiểm soát sự tăng sinh tế bào và kích thước cơ quan. Do sự biểu hiện *CYCLIN D3;1-10* và sự tăng sinh tế bào tăng lên, các đột biến mất chức năng *GmKIX8-1* đã biểu hiện sự gia tăng rõ rệt về kích thước của lá và hạt (Nguyen và cs, 2021). Rất gần đây, ở cả *Arabidopsis* và đậu tương, một tầng điều tiết quan trọng liên quan đến CO (chất điều hòa trung tâm của con đường quang kỳ) và AP2 (đặc điểm nhận dạng mô phân sinh hoa) đã được chứng minh là trung gian hóa kích thước hạt được điều hòa bởi quang kỳ theo cách phụ thuộc vào mẹ (Yu và cs, 2023). *GmCOL2b* (một chất tương đồng CO của đậu tương) thúc đẩy tích cực kích thước hạt trong thời gian ngắn ngày bằng cách ức chế trực tiếp sự biểu hiện của *GmAP2-1* và *GmAP2-2*.

4. Gen quy định dầu hạt

Hạt giống tích lũy dự trữ, bao gồm dầu, protein và tinh bột, được lấp đầy trong quá trình phát triển và trưởng thành của hạt. Do đó, hiểu được chất lưu trữ được nạp vào hạt là rất quan trọng để cải thiện năng suất cây trồng và chất lượng dinh dưỡng. Trong những thập kỷ qua, nhiều nỗ lực đã được thực hiện đối với việc mổ xẻ các con đường phân tử để tích lũy dự trữ hạt giống, đặc biệt là ở cây *Arabidopsis*. Các TF, chẳng hạn như LEC1, LEC2, ABI3, FUS3 và WRI1, và các chất kích hoạt hoặc chất ức chế khác để tích lũy dự trữ trong quá trình phát triển hạt, đã được xác định ở thực vật (Yang và cs, 2022a). Tuy nhiên, nhiều chi tiết và cơ chế vẫn chưa được làm rõ, đặc biệt là đối với các loại cây trồng thiết yếu như đậu tương (Hình 3; Bảng 1).

LEC1 là một tiểu đơn vị TF không điển hình (NF-YB) tương tác với các tiểu đơn vị NF-YA và NF-YC để tạo thành phức hợp NF-Y TF. Nó là trung tâm để kiểm soát sự phát triển của hạt giống, chẳng hạn như hình thái phôi, phát triển nội nhũ và tích lũy dự trữ (Jo và cs, 2019). Ở cây *Arabidopsis*, các đột biến *lec1* không có giá trị thể hiện những khiếm khuyết nổi bật trong phôi và hạn chế nghiêm trọng sự tích lũy protein và lipid trong hạt

(Meinke và cs, 1994; West và cs, 1994). Hơn nữa, sự biểu hiện quá mức của *LEC1* đã gây ra sự kích hoạt các gen liên quan đến sự tích lũy protein và lipid dự trữ, dẫn đến tăng hàm lượng lipid và FA trong cây *Arabidopsis* chuyển gen (Kagaya và cs, 2005). Ở đậu tương, *GmLEC1* (*GmLEC1a* hoặc *GmLEC1b*) điều hòa phiên mã các gen liên quan đến các quá trình tế bào riêng biệt trong quá trình phát triển hạt và kích hoạt quá trình sinh tổng hợp FA của hạt (Pelletier và cs, 2017; Zhang và cs, 2017). Nghiên cứu sâu hơn cho thấy rằng *GmLEC1* hoạt động kết hợp với các TF như *GmAREB3*, *GmbZIP67* và *GmABI3* để điều chỉnh sự phát triển của hạt đậu tương (Jo và cs, 2020).

LEC1 tương tác vật lý với LEC2, TF miền liên kết DNA B3, có vai trò điều tiết quan trọng trong quá trình phát triển hạt giống và kiểm soát mức độ protein và dầu hạt trong cây *Arabidopsis* (Santos-Mendoza và cs, 2008; Angeles-Núñez và Tiessen, 2011; Kim và cs, 2015; Jo và cs, 2019). Các hạt đột biến mất chức năng *lec2* cho thấy lượng dầu và protein giảm lần lượt là 30% và 15%, nhưng tích lũy nhiều tinh bột và sucrose hơn so với các loại hạt hoang dại (Angeles-Núñez và Tiessen, 2011). Ngược lại, ở cả cây *Arabidopsis* và cây thuốc lá biến đổi gen, biểu hiện cảm ứng *AtLEC2* làm tăng tích lũy dự trữ dầu, chẳng hạn như TAG và FA (Mendoza và cs, 2005; Andrianov và cs, 2010; Kim và cs, 2015). Ở đậu tương, *GmLEC2* quy định một tập hợp con các gen liên quan đến quá trình chuyển hóa chất dự trữ hạt giống (Manan và cs, 2017). So với hạt đối chứng, hàm lượng TAG và FA chuỗi dài của hạt *Arabidopsis* chuyển gen biểu hiện quá mức *GmLEC2a* tăng lần lượt là 34% và 4%.

Trong mạng lưới phiên mã của quá trình tích lũy dầu hạt ở cây *Arabidopsis*, LEC1 và LEC2 cùng nhau thúc đẩy sự biểu hiện WRI1, một gen AP2 TF chịu trách nhiệm điều hòa phiên mã quá trình sinh tổng hợp dầu và cơ chế điều hòa này được bảo tồn ở các loài thực vật khác, chẳng hạn như đậu tương và ngô (Baud và cs, 2007; Mu và cs, 2008; Shen và cs, 2010; Manan và cs, 2017; Pelletier và cs, 2017; Yang và cs, 2022a). Hai ortholog đậu tương của nó, *GmWRI1a* và *GmWRI1b*, đóng vai trò trung tâm trong việc tích lũy dầu hạt. Sự biểu hiện quá mức của *GmWRI1a* hoặc *GmWRI1b* làm tăng đáng kể tổng lượng dầu và hàm lượng FA, đồng thời thay đổi thành phần FA trong hạt, trong khi *GmWRI1* loại bỏ rễ tơ lại cản trở quá trình sinh tổng hợp lipid (Chen và cs, 2018; Chen và cs, 2020; Guo và cs. , 2020; Wang và cs, 2022).

GmZF392, một protein nhánh kẽm CCCH (TZF) song song đặc trưng cho hạt giống, thúc đẩy sự tích lũy dầu hạt bằng cách nhắm mục tiêu một yếu tố cis lưỡng cực với các trình tự giàu TA và TG trong các vùng khởi động, do đó kích hoạt biểu hiện gen xuôi dòng liên quan đến quá trình sinh tổng hợp lipid (Lu và cs, 2021). *GmZF392* tương tác vật lý với *GmZF351*, một chất kích hoạt khác của sự tích lũy lipid, để bổ sung/hiệp đồng làm tăng biểu hiện của các gen sinh tổng hợp lipid xuôi dòng (Li và cs, 2017; Lu và cs, 2021). Và cả *GmZF392* và *GmZF351* đều được điều chỉnh tích cực bởi *GmNFYA*, một TF tương quan với hàm lượng dầu (Lu và cs, 2016; Lu và cs, 2021). Ngoài ra, *GmZF392* và *GmZF351* cũng là mục tiêu trực tiếp của *GmLEC1* (Pelletier và cs, 2017). Quan trọng hơn, *GmZF392* và *GmZF351* đã được chọn lọc bằng cách thuần hóa từ đậu tương dại sang đậu tương trồng.

Ngoài các TF ở trên tạo thành mô-đun điều hòa, một số gen chức năng cũng tham gia vào việc điều chỉnh hàm lượng dầu hạt ở đậu tương (Hình 3; Bảng 1). Sự biểu hiện quá mức của gen *bZIP TF* (*GmbZIP123*) giúp tăng cường tích lũy lipid trong hạt *Arabidopsis* biến đổi gen thông qua điều hòa vận chuyển đường (Song và cs, 2013). *GmB1*, mã hóa protein xuyên màng giống như chất vận chuyển để sinh tổng hợp phấn ở vỏ quả, không chỉ kiểm soát phấn của vỏ hạt ở đậu tương dại mà còn ảnh hưởng đến hàm lượng dầu ở đậu tương

trồng (Zhang và cs, 2018a). *GmOLEO1*, một gen mã hóa protein oleosin được chọn lọc nhân tạo mạnh, giúp cải thiện hàm lượng dầu hạt trong quá trình thuần hóa đậu tương bằng cách ảnh hưởng đến quá trình chuyển hóa TAG (Zhang và cộng sự, 2019b).

5. Gen quy định protein trong hạt

So với kích thước hạt và hàm lượng dầu, chỉ có một số gen kiểm soát hàm lượng protein hoặc axit amin của hạt đã được xác định chức năng (Hình 3; Bảng 1) (Krishnan và Jez, 2018). *GTPase GmRab5a* nhỏ và các yếu tố trao đổi guanine *GmVPS9* của nó được chứng minh là hoạt động trong quá trình tích lũy protein dự trữ sau Golgi ở đậu tương (Wei và cs, 2020). Sự biểu hiện quá mức thoáng qua của biến thể âm tính chiếm ưu thế của *GmRab5a*, hoặc RNAi của *GmRab5a* hoặc *GmVPS9s*, rõ ràng đã làm giảm quá trình vận chuyển chất đánh dấu hàng hóa, vốn được sử dụng để phản ánh quá trình vận chuyển protein dự trữ đến các không bào dự trữ protein trong các tế bào lá mầm đậu tương. Ngoài ra, một số gen, bao gồm *POWR1*, *GmSWEET10a*, *GmSWEET10b* và *GmST05*, điều chỉnh đa hướng protein hạt, hàm lượng dầu và kích thước hạt (Wang và cs, 2020b; Duan và cs, 2022; Goettel và cs, 2022), điều đó được thảo luận chi tiết trong phần tiếp theo.

6. Gen quy định kích thước hạt, hàm lượng dầu, protein của hạt

Kích thước hạt, tích lũy dầu và hàm lượng protein trong đậu tương là những đặc điểm nông học có mối tương quan cao. Tuy nhiên, việc lựa chọn và cơ sở phân tử cơ bản của các đặc điểm tương quan với hạt giống này trong quá trình thuần hóa đậu tương vẫn chưa được hiểu rõ, đây là một trong những trở ngại đối với việc cải thiện năng suất và chất lượng đậu tương. Cho đến nay, một số gen điều hòa đa hướng kiểm soát kích thước hạt, tích lũy dầu và hàm lượng protein đã được nhân bản và xác định chức năng trong đậu tương (Hình 3; Bảng 1).

Ví dụ, sự biểu hiện lệch vị trí của *GmDof4*, *GmDof11*, *GmMYB73* và *GmDREBL* đã tăng cường cả kích thước/trọng lượng hạt và sự tích lũy dầu trong hạt *Arabidopsis* chuyển gen (Wang và cs, 2007; Liu và cs, 2014; Zhang và cs, 2016b). *GmPDAT*, một gen mã hóa phospholipid diacylglycerol acyltransferase, được biểu hiện cao hơn ở các giống đậu tương hạt to và nhiều dầu so với các giống đậu tương hạt nhỏ và ít dầu. Sự biểu hiện quá mức của *GmPDAT* đã cải thiện kích thước hạt và mức dầu, trong khi thực vật *GmPDAT* RNAi đã giảm kích thước hạt và tích lũy dầu (Liu và cs, 2020a). *GmST1* mã hóa UDP-D-glucuronate 4-epimerase điều chỉnh tích cực kích thước hạt và hàm lượng dầu bằng cách điều chỉnh quá trình sinh tổng hợp pectin và đường phân, đồng thời trải qua quá trình chọn lọc trong quá trình thuần hóa đậu tương (Li và cs, 2022a).

Các thành viên họ *SWEET* vận chuyển đường đóng vai trò quan trọng trong quá trình phát triển hạt giống (Chen và cs, 2015; Wang và cs, 2019). Một cặp gen *SWEET* parolog trong đậu tương, *GmSWEET10a* và *GmSWEET10b*, đã trải qua quá trình chọn lọc từng bước làm thay đổi đồng bộ kích thước hạt, sự tích lũy dầu và mức protein trong quá trình thuần hóa đậu tương, bằng cách điều chỉnh quá trình phân loại đường từ vỏ hạt đến phôi (Zhang và cs, 2020; Wang và cs, 2020b). So với cây dại, cây đậu tương biểu hiện *GmSWEET10a* hoặc *GmSWEET10b* có kích thước hạt tăng đáng kể và lượng dầu tích lũy cao hơn nhưng lại giảm mức protein, trong khi các cây loại bỏ của chúng đã giảm kích thước hạt và hàm lượng dầu nhưng tăng mức độ protein (Wang và cs, 2020b). Rất gần đây, một thành viên họ protein gắn phosphatidylethanolamine (PEBP), *GmST05* (còn được gọi là *GmMFT*), đã được chứng minh là điều chỉnh tích cực kích thước hạt cũng như thay đổi mức dầu và protein, có thể bằng cách ảnh hưởng đến quá trình phiên mã

GmSWEET10a (Li và cs, 2014; Duan và cs, 2022). Ngoài ra, một gen miền CCT, *POWR1*, được thuần hóa và điều chỉnh về mặt đa hướng chất lượng và năng suất hạt giống ở đậu tương, có thể bằng cách điều chỉnh quá trình chuyển hóa lipid và vận chuyển chất dinh dưỡng (Goettel và cs, 2022). Việc chèn phần tử có thể thay thế (TE) vào miền CCT của *POWR1* dẫn đến tăng trọng lượng hạt và hàm lượng dầu nhưng giảm hàm lượng protein. Ngược lại, sự biểu hiện quá mức của *POWR1* cho thấy hàm lượng protein được cải thiện và làm giảm trọng lượng hạt cũng như tích lũy dầu ở cây chuyển gen.

7. Thách thức và triển vọng

Kích thước hạt, hàm lượng dầu và protein là những tính trạng số lượng phức tạp do nhiều gen chi phối. Mặc dù lập bản đồ liên kết và phân tích GWAS đã xác định được nhiều QTL kiểm soát kích thước hạt, tích lũy dầu và hàm lượng protein trong đậu tương, nhưng chỉ có một số gen được phân lập và xác nhận chức năng. Một lý do cơ bản cho hiện tượng này là các nhà nghiên cứu này thường chỉ sử dụng một hoặc hai cách tiếp cận, khiến cho việc xác định mục tiêu tiềm ẩn những đặc điểm hạt giống này trở nên khó khăn. Một trở ngại quan trọng khác là thiếu một hệ thống chuyển đổi gen đậu tương nhanh và hiệu quả cho các kiểu gen đậu tương khác nhau, chẳng hạn như chuyển đổi nút lá mầm qua *Agrobacterium* trung gian, đã được sử dụng rộng rãi trong những năm gần đây. Hệ thống chuyển đổi gen chậm và không hiệu quả khiến việc xác định và xác minh chức năng của gen đậu tương trở nên khó khăn hơn (Zhang và cs, 2022). Đó là lý do tại sao, trong một số nghiên cứu, đặc biệt là những nghiên cứu trước năm 2015, việc xác nhận chức năng đã được thực hiện trên cây *Arabidopsis* thay vì cây đậu tương.

Với sự tiến bộ nhanh chóng của nghiên cứu omics và giảm chi phí thử nghiệm, ngày càng có nhiều dữ liệu về omics đậu tương được tạo ra, chẳng hạn như bộ gen sắp xếp lại trình tự, bộ phiên mã, bộ chuyển hóa, bộ gen proteome, bộ gen biểu sinh, bộ gen pan-gen và bộ gen 3D (Ohyanagi và cs, 2012; Lin và cs, 2014; Shen và cs, 2014; Zhou và cs, 2015; Liu và cs, 2016; Fang và cs, 2017; Shen và cs, 2018a; Shen và cs, 2018b; Liu và cs, 2020b; Silva và cs, 2021; Ni và cs, 2023). Những tài nguyên omics được phát hành này sẽ thúc đẩy rộng rãi nghiên cứu về bộ gen chức năng của đậu tương. Hiện tại, như GWAS, TWAS (nghiên cứu liên kết trên toàn hệ phiên mã), EWAS (nghiên cứu liên kết trên toàn hệ biểu mô) và PWAS (nghiên cứu liên kết trên toàn hệ protein), cũng như các nghiên cứu liên kết dữ liệu đa omics, chẳng hạn như eGWAS (biểu hiện gen- nghiên cứu liên kết trên toàn bộ bộ gen) và mGWAS (nghiên cứu liên kết trên toàn bộ bộ gen dựa trên chuyển hóa) đã được phát triển và ứng dụng thành công (Shen và cs, 2022). Việc tích hợp nhiều cách tiếp cận omics sẽ cung cấp thêm manh mối và giúp thu hẹp phạm vi mục tiêu nằm bên dưới những đặc điểm hạt giống này. Tuy nhiên, việc sử dụng những dữ liệu omics khổng lồ này tồn tại ở nhiều dạng khác nhau là một thách thức đáng kể. Do đó, các phương pháp toán học, chẳng hạn như phân tích tổng hợp, được kỳ vọng sẽ giải quyết vấn đề đó. Ngoài ra, công nghệ trí tuệ nhân tạo (AI) hoặc phương pháp tiếp cận máy móc có thể giúp việc khai thác dữ liệu lớn hiệu quả hơn, chẳng hạn như xử lý dữ liệu omics, xây dựng cấu trúc protein và tích hợp dữ liệu pan-omics (Baek và cs, 2021; Jumper và cs, 2021; Reel và cs, 2021).

Công nghệ chỉnh sửa bộ gen dựa trên CRISPR/Cas cho phép sửa đổi chính xác bộ gen để thu được các tính trạng mong muốn và có thể dự đoán được đã được áp dụng thành công cho nghiên cứu chức năng gen và tạo tế bào mầm cây trồng. So với các loại cây trồng khác, chẳng hạn như lúa gạo, quá trình chỉnh sửa bộ gen đậu tương chủ yếu ở giai đoạn sơ khai; tuy nhiên, những câu chuyện thành công đã chứng minh tính khả thi của việc chỉnh sửa gen ở đậu tương (Cai và cs, 2018; Bai và cs, 2020; Wang và cs, 2020b; Nguyen

và cs, 2021; Bai và cs, 2022; Duan và cs, 2022; Hu và cs, 2022; Liang và cs, 2022; Li và cs, 2022a). Trong tương lai, quá trình biến đổi đậu tương được cải thiện và nhiều ứng dụng hơn của "chỉnh sửa cơ sở" đơn gen hoặc đa gen sẽ tạo điều kiện thuận lợi cho nghiên cứu chức năng ở đậu tương, cuối cùng cho phép chúng ta giải mã các đặc điểm hạt phức tạp này và xác định các gen quan trọng làm cơ sở cho kích thước hạt, hàm lượng dầu và protein.

Mục tiêu cuối cùng của chọn giống đậu tương là trồng đậu tương có năng suất và chất lượng cao. Cho đến nay, chọn giống cây trồng đã phát triển từ chọn lọc nhân tạo (giai đoạn 1.0) và chọn giống lai (giai đoạn 2.0) đến chọn giống phân tử (giai đoạn 3.0). Tuy nhiên, để giải quyết cuộc khủng hoảng thiếu lương thực do dân số ngày càng tăng gây ra, việc chọn giống thông minh (giai đoạn 4.0) có thể nhanh chóng tổng hợp các alen xuất sắc thông qua thiết kế chính xác sẽ xuất hiện (Shen và cs, 2022). Trong các giai đoạn chọn giống trước đây, các nhà lai tạo thường phải xếp các tính trạng mong muốn vào một dòng duy nhất để tạo ra một siêu giống, đây là một nhiệm vụ rất lớn. Trong giai đoạn chọn giống 4.0, thiết kế tối ưu và chính xác để nhanh chóng hình thành tháp nhiều alen ưu tú với các đặc điểm hạt giống mong muốn sẽ tạo điều kiện cải thiện năng suất, dầu và hàm lượng protein trong đậu tương.

Tài liệu tham khảo

Andrianov, V., Borisjuk, N., Pogrebnyak, N., Brinker, A., Dixon, J., Spitsin, S., et al. (2010). Tobacco as a production platform for biofuel: Overexpression of *Arabidopsis DGAT* and *LEC2* genes increases accumulation and shifts the composition of lipids in green biomass. *Plant Biotechnol. J.* 8 (3), 277–287. doi: 10.1111/j.1467-7652.2009.00458.x

PubMed Abstract | CrossRef Full Text | Google Scholar

Angeles-Núñez, J. G., Tiessen, A. (2011). Mutation of the transcription factor LEAFY COTYLEDON 2 alters the chemical composition of *Arabidopsis* seeds, decreasing oil and protein content, while maintaining high levels of starch and sucrose in mature seeds. *J. Plant Physiol.* 168 (16), 1891–1900. doi: 10.1016/j.jplph.2011.05.003

PubMed Abstract | CrossRef Full Text | Google Scholar

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373 (6557), 871–876. doi: 10.1126/science.abj8754

PubMed Abstract | CrossRef Full Text | Google Scholar

Baekelandt, A., Pauwels, L., Wang, Z., Li, N., De Milde, L., Natran, A., et al. (2018). *Arabidopsis* leaf flatness is regulated by PPD2 and NINJA through repression of *CYCLIN D3* genes. *Plant Physiol.* 178 (1), 217–232. doi: 10.1104/pp.18.00327

PubMed Abstract | CrossRef Full Text | Google Scholar

Bai, M., Yuan, J., Kuang, H., Gong, P., Li, S., Zhang, Z., et al. (2020). Generation of a multiplex mutagenesis population *via* pooled CRISPR-Cas9 in soybean. *Plant Biotechnol. J.* 18 (3), 721–731. doi: 10.1111/pbi.13239

PubMed Abstract | CrossRef Full Text | Google Scholar

Bai, M., Yuan, C., Kuang, H., Sun, Q., Hu, X., Cui, L., et al. (2022). Combination of two multiplex genome-edited soybean varieties enables customization of protein functional properties. *Mol. Plant* 15 (7), 1081–1083. doi: 10.1016/j.molp.2022.05.011

PubMed Abstract | CrossRef Full Text | Google Scholar

Bates, P. D., Stymne, S., Ohlrogge, J. (2013). Biochemical pathways in seed oil synthesis. *Curr. Opin. Plant Biol.* 16 (3), 358–364. doi: 10.1016/j.pbi.2013.02.015

PubMed Abstract | CrossRef Full Text | Google Scholar

Baud, S., Mendoza, M. S., To, A., Harscoët, E., Lepiniec, L., Dubreucq, B. (2007). WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON 2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. *Plant J.* 50 (5), 825–838. doi: 10.1111/j.1365-313X.2007.03092.x

PubMed Abstract | CrossRef Full Text | Google Scholar

Cai, Y., Chen, L., Liu, X., Guo, C., Sun, S., Wu, C., et al. (2018). CRISPR/Cas9-mediated targeted mutagenesis of *GmFT2a* delays flowering time in soya bean. *Plant Biotechnol. J.* 16 (1), 176–185. doi: 10.1111/pbi.12758

PubMed Abstract | CrossRef Full Text | Google Scholar

Cao, Y., Li, S., Wang, Z., Chang, F., Kong, J., Gai, J., et al. (2017). Identification of major quantitative trait loci for seed oil content in soybeans by combining linkage and genome-wide association mapping. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01222

CrossRef Full Text | Google Scholar

Carter, T. E., Jr., Nelson, R. L., Sneller, C. H., Cui, Z. (2004). Genetic diversity in soybean. *Soybeans: Improvement production uses* 16, 303–416. doi: 10.2134/agronmonogr16.3ed.c8

CrossRef Full Text | Google Scholar

Chen, L., Lin, I., Qu, X., Sosso, D., McFarlane, H. E., Londoño, A., et al. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the *Arabidopsis* embryo. *Plant Cell* 27 (3), 607–619. doi: 10.1105/tpc.114.134585

PubMed Abstract | CrossRef Full Text | Google Scholar

Chen, B., Zhang, G., Li, P., Yang, J., Guo, L., Benning, C., et al. (2020). Multiple *GmWRI1s* are redundantly involved in seed filling and nodulation by regulating plastidic glycolysis, lipid biosynthesis and hormone signalling in soybean (*Glycine max*). *Plant Biotechnol. J.* 18 (1), 155–171. doi: 10.1111/pbi.13183

PubMed Abstract | CrossRef Full Text | Google Scholar

Chen, L., Zheng, Y., Dong, Z., Meng, F., Sun, X., Fan, X., et al. (2018). Soybean (*Glycine max*) WRINKLED1 transcription factor, GmWRI1a, positively regulates seed oil accumulation. *Mol. Genet. Genomics* 293 (2), 401–415. doi: 10.1007/s00438-017-1393-2

PubMed Abstract | CrossRef Full Text | Google Scholar

Cui, B., Chen, L., Yang, Y., Liao, H. (2020). Genetic analysis and map-based delimitation of a major locus *qSS3* for seed size in soybean. *Plant Breed.* 139 (6), 1145–1157. doi: 10.1111/pbr.12853

CrossRef Full Text | Google Scholar

Du, J., Wang, S., He, C., Zhou, B., Ruan, Y. L., Shou, H. (2017). Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J. Exp. Bot.* 68 (8), 1955–1972. doi: 10.1093/jxb/erw460

PubMed Abstract | CrossRef Full Text | Google Scholar

Duan, Z., Zhang, M., Zhang, Z., Liang, S., Fan, L., Yang, X., et al. (2022). Natural allelic variation of *GmST05* controlling seed size and quality in soybean. *Plant Biotechnol. J.* 20 (9), 1807–1818. doi: 10.1111/pbi.13865

PubMed Abstract | CrossRef Full Text | Google Scholar

Eskandari, M., Cober, E. R., Rajcan, I. (2013a). Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. *Theor. Appl. Genet.* 126 (2), 483–495. doi: 10.1007/s00122-012-1995-3

PubMed Abstract | CrossRef Full Text | Google Scholar

Eskandari, M., Cober, E. R., Rajcan, I. (2013b). Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. *Theor. Appl. Genet.* 126 (6), 1677–1687. doi: 10.1007/s00122-013-2083-z

PubMed Abstract | CrossRef Full Text | Google Scholar

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 18 (1), 161. doi: 10.1186/s13059-017-1289-9

PubMed Abstract | CrossRef Full Text | Google Scholar

Fang, W., Wang, Z., Cui, R., Li, J., Li, Y. (2012). Maternal control of seed size by *EOD3/CYP78A6* in *Arabidopsis thaliana*. *Plant J.* 70 (6), 929–939. doi: 10.1111/j.1365-313X.2012.04907.x

PubMed Abstract | CrossRef Full Text | Google Scholar

Fliege, C. E., Ward, R. A., Vogel, P., Nguyen, H., Quach, T., Guo, M., et al. (2022). Fine mapping and cloning of the major seed protein quantitative trait loci on soybean chromosome 20. *Plant J.* 110 (1), 114–128. doi: 10.1111/tpj.15658

PubMed Abstract | CrossRef Full Text | Google Scholar

Ge, L., Yu, J., Wang, H., Luth, D., Bai, G., Wang, K., et al. (2016). Increasing seed size and quality by manipulating *BIG SEEDS1* in legume species. *Proc. Natl. Acad. Sci. U. S. A.* 113 (44), 12414–12419. doi: 10.1073/pnas.1611763113

PubMed Abstract | CrossRef Full Text | Google Scholar

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: the challenge of feeding 9 billion people. *Science* 327 (5967), 812–818. doi: 10.1126/science.1185383

PubMed Abstract | CrossRef Full Text | Google Scholar

Goettel, W., Zhang, H., Li, Y., Qiao, Z., Jiang, H., Hou, D., et al. (2022). *POWR1* is a domestication gene pleiotropically regulating seed quality and yield in soybean. *Nat. Commun.* 13 (1), 3051. doi: 10.1038/s41467-022-30314-7

PubMed Abstract | CrossRef Full Text | Google Scholar

Gu, Y., Li, W., Jiang, H., Wang, Y., Gao, H., Liu, M., et al. (2017). Differential expression of a *WRKY* gene between wild and cultivated soybeans correlates to seed size. *J. Exp. Bot.* 68 (11), 2717–2729. doi: 10.1093/jxb/erx147

PubMed Abstract | CrossRef Full Text | Google Scholar

Guo, W., Chen, L., Chen, H., Yang, H., You, Q., Bao, A., et al. (2020). Overexpression of *GmWRI1b* in soybean stably improves plant architecture and associated yield

parameters, and increases total seed oil production under field conditions. *Plant Biotechnol. J.* 18 (8), 1639–1641. doi: 10.1111/pbi.13324

PubMed Abstract | CrossRef Full Text | Google Scholar

Han, Y., Li, D., Zhu, D., Li, H., Li, X., Teng, W., et al. (2012). QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theor. Appl. Genet.* 125 (4), 671–683. doi: 10.1007/s00122-012-1859-x

PubMed Abstract | CrossRef Full Text | Google Scholar

He, Q., Xiang, S., Yang, H., Wang, W., Shu, Y., Li, Z., et al. (2021). A genome-wide association study of seed size, protein content, and oil content using a natural population of sichuan and chongqing soybean. *Euphytica* 217(11), 198. doi: 10.1007/s10681-021-02931-8

CrossRef Full Text | Google Scholar

Hong, H., Najafabadi, M. Y., Torkamaneh, D., Rajcan, I. (2022). Identification of quantitative trait loci associated with seed quality traits between Canadian and Ukrainian mega-environments using genome-wide association study. *Theor. Appl. Genet.* 135 (7), 2515–2530. doi: 10.1007/s00122-022-04134-8

PubMed Abstract | CrossRef Full Text | Google Scholar

Hu, D., Li, X., Yang, Z., Liu, S., Hao, D., Chao, M., et al. (2022). Downregulation of a gibberellin 3β-hydroxylase enhances photosynthesis and increases seed yield in soybean. *New Phytol.* 235 (2), 502–517. doi: 10.1111/nph.18153

PubMed Abstract | CrossRef Full Text | 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. doi: 10.1186/1471-2164-15-1

PubMed Abstract | CrossRef Full Text | Google Scholar

Jo, L., Pelletier, J. M., Harada, J. J. (2019). Central role of the LEAFY COTYLEDON1 transcription factor in seed development. *J. Integr. Plant Biol.* 61 (5), 564–580. doi: 10.1111/jipb.12806

PubMed Abstract | CrossRef Full Text | Google Scholar

Jo, L., Pelletier, J. M., Hsu, S. W., Baden, R., Goldberg, R. B., Harada, J. J. (2020). Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. *Proc. Natl. Acad. Sci. U. S. A.* 117 (2), 1223–1232. doi: 10.1073/pnas.1918441117

PubMed Abstract | CrossRef Full Text | Google Scholar

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi: 10.1038/s41586-021-03819-2

PubMed Abstract | CrossRef Full Text | Google Scholar

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., Hattori, T. (2005). LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of *FUSCA3* and *ABSCISIC ACID INSENSITIVE3*. *Plant Cell Physiol.* 46 (3), 399–406. doi: 10.1093/pcp/pci048

PubMed Abstract | CrossRef Full Text | Google Scholar

Kim, H. U., Lee, K. R., Jung, S. J., Shin, H. A., Go, Y. S., Suh, M. C., et al. (2015). Senescence-inducible *LEC2* enhances triacylglycerol accumulation in leaves without negatively affecting plant growth. *Plant Biotechnol. J.* 13 (9), 1346–1359. doi: 10.1111/pbi.12354

PubMed Abstract | CrossRef Full Text | Google Scholar

Krishnan, H. B., Jez, J. M. (2018). Review: The promise and limits for enhancing sulfur-containing amino acid content of soybean seed. *Plant Sci.* 272, 14–21. doi: 10.1016/j.plantsci.2018.03.030

PubMed Abstract | CrossRef Full Text | Google Scholar

Kumar, R., Saini, M., Taku, M., Debbarma, P., Mahto, R. K., Ramlal, A., et al. (2022). Identification of quantitative trait loci (QTLs) and candidate genes for seed shape and 100-seed weight in soybean [Glycine max (L.) merr.]. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1074245

CrossRef Full Text | Google Scholar

Kumawat, G., Xu, D. (2021). A major and stable quantitative trait locus *qSS2* for seed size and shape traits in a soybean RIL population. *Front. Genet.* 12. doi: 10.3389/fgene.2021.646102

PubMed Abstract | CrossRef Full Text | Google Scholar

Lee, S., Van, K., Sung, M., Nelson, R., LaMantia, J., McHale, L. K., et al. (2019). Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. *Theor. Appl. Genet.* 132(6), 1639–1659. doi: 10.1007/s00122-019-03304-5

PubMed Abstract | CrossRef Full Text | Google Scholar

Li, Q., Fan, C., Zhang, X., Wang, X., Wu, F., Hu, R., et al. (2014). Identification of a soybean *MOTHER OF FT AND TFL1* homolog involved in regulation of seed germination. *PloS One* 9 (6), e99642. doi: 10.1371/journal.pone.0099642

PubMed Abstract | CrossRef Full Text | Google Scholar

Li, Q., Lu, X., Song, Q., Chen, H., Wei, W., Tao, J., et al. (2017). Selection for a zinc-finger protein contributes to seed oil increase during soybean domestication. *Plant Physiol.* 173 (4), 2208–2224. doi: 10.1104/pp.16.01610

PubMed Abstract | CrossRef Full Text | Google Scholar

Li, Q., Lu, X., Wang, C., Shen, L., Dai, L., He, J., et al. (2022b). Genome-wide association study and transcriptome analysis reveal new QTL and candidate genes for nitrogen-deficiency tolerance in rice. *Crop J.* 10 (4), 942–951. doi: 10.1016/j.cj.2021.12.006

CrossRef Full Text | Google Scholar

Li, N., Xu, R., Li, Y. (2019). Molecular networks of seed size control in plants. *Annu. Rev. Plant Biol.* 70, 435–463. doi: 10.1146/annurev-arplant-050718-095851

PubMed Abstract | CrossRef Full Text | Google Scholar

Li, J., Zhang, Y., Ma, R., Huang, W., Hou, J., Fang, C., et al. (2022a). Identification of *ST1* reveals a selection involving hitchhiking of seed morphology and oil content during soybean domestication. *Plant Biotechnol. J.* 20 (6), 1110–1121. doi: 10.1111/pbi.13791

PubMed Abstract | CrossRef Full Text | Google Scholar

Liang, Q., Chen, L., Yang, X., Yang, H., Liu, S., Kou, K., et al. (2022). Natural variation of *Dt2* determines branching in soybean. *Nat. Commun.* 13 (1), 6429. doi: 10.1038/s41467-022-34153-4

PubMed Abstract | CrossRef Full Text | Google Scholar

Lin, H., Rao, J., Shi, J., Hu, C., Cheng, F., Wilson, Z. A., et al. (2014). Seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. *J. Integr. Plant Biol.* 56 (9), 826–836. doi: 10.1111/jipb.12228

PubMed Abstract | CrossRef Full Text | Google Scholar

Liu, K. (1997). "Chemistry and nutritional value of soybean components," in *Soybean: Chemistry, technology and utilization*. (Boston, MA: Springer), 25–113. doi: 10.1007/978-1-4615-1763-4_2

CrossRef Full Text | Google Scholar

Liu, A., Cheng, S., Yung, W., Li, M., Lam, H. (2022). Genetic regulations of the oil and protein contents in soybean seeds and strategies for improvement. *Adv. Bot. Res*. 102, 259–293. doi: 10.1016/bs.abr.2022.03.002

CrossRef Full Text | Google Scholar

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., et al. (2020b). Pan-genome of wild and cultivated soybeans. *Cell* 182 (1), 162–176.e113. doi: 10.1016/j.cell.2020.05.023

PubMed Abstract | CrossRef Full Text | Google Scholar

Liu, T., Fang, C., Ma, Y., Shen, Y., Li, C., Li, Q., et al. (2016). Global investigation of the co-evolution of MIRNA genes and micro RNA targets during soybean domestication. *Plant J.* 85 (3), 396–409. doi: 10.1111/tpj.13113

PubMed Abstract | CrossRef Full Text | Google Scholar

Liu, Y., Li, Q., Lu, X., Song, Q., Lam, S. M., Zhang, W., et al. (2014). Soybean *GmMYB73* promotes lipid accumulation in transgenic plants. *BMC Plant Biol.* 14(1), 73. doi: 10.1186/1471-2229-14-73

PubMed Abstract | CrossRef Full Text | Google Scholar

Liu, J., Zhang, Y., Han, X., Zuo, J., Zhang, Z., Shang, H., et al. (2020a). An evolutionary population structure model reveals pleiotropic effects of *GmPDAT* for traits related to seed size and oil content in soybean. *J. Exp. Bot.* 71 (22), 6988–7002. doi: 10.1093/jxb/eraa426

PubMed Abstract | CrossRef Full Text | Google Scholar

Lu, X., Li, Q., Xiong, Q., Li, W., Bi, Y., Lai, Y., et al. (2016). The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *Plant J.* 86 (6), 530–544. doi: 10.1111/tpj.13181

PubMed Abstract | CrossRef Full Text | Google Scholar

Lu, L., Wei, W., Li, Q. T., Bian, X. H., Lu, X., Hu, Y., et al. (2021). A transcriptional regulatory module controls lipid accumulation in soybean. *New Phytol.* 231 (2), 661–678. doi: 10.1111/nph.17401

PubMed Abstract | CrossRef Full Text | Google Scholar

Lu, X., Xiong, Q., Cheng, T., Li, Q. T., Liu, X. L., Bi, Y. D., et al. (2017). A *PP2C-1* allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* 10 (5), 670–684. doi: 10.1016/j.molp.2017.03.006

PubMed Abstract | CrossRef Full Text | Google Scholar

Luo, M., Dennis, E. S., Berger, F., Peacock, W. J., Chaudhury, A. (2005). *MINISEED3* (*MINI3*), a WRKY family gene, and *HAIKU2* (*IKU2*), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 102 (48), 17531–17536. doi: 10.1073/pnas.0508418102

PubMed Abstract | CrossRef Full Text | Google Scholar

Luo, S., Jia, J., Liu, R., Wei, R., Guo, Z., Cai, Z., et al. (2022). Identification of major QTLs for soybean seed size and seed weight traits using a RIL population in different environments. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1094112

CrossRef Full Text | Google Scholar

Manan, S., Ahmad, M. Z., Zhang, G., Chen, B., Haq, B. U., Yang, J., et al. (2017). Soybean LEC2 regulates subsets of genes involved in controlling the biosynthesis and catabolism of seed storage substances and seed development. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01604

PubMed Abstract | CrossRef Full Text | Google Scholar

Meinke, D. W., Franzmann, L. H., Nickle, T. C., Yeung, E. C. (1994). Leafy cotyledon mutants of *Arabidopsis*. *Plant Cell* 6 (8), 1049–1064. doi: 10.1105/tpc.6.8.1049

PubMed Abstract | CrossRef Full Text | Google Scholar

Mendoza, M. S., Dubreucq, B., Miquel, M., Caboche, M., Lepiniec, L. (2005). LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Lett.* 579 (21), 4666–4670. doi: 10.1016/j.febslet.2005.07.037

PubMed Abstract | CrossRef Full Text | Google Scholar

Miao, L., Yang, S., Zhang, K., He, J., Wu, C., Ren, Y., et al. (2020). Natural variation and selection in *GmSWEET39* affect soybean seed oil content. *New Phytol.* 225 (4), 1651–1666. doi: 10.1111/nph.16250

PubMed Abstract | CrossRef Full Text | Google Scholar

Mu, J., Tan, H., Zheng, Q., Fu, F., Liang, Y., Zhang, J., et al. (2008). LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in *Arabidopsis*. *Plant Physiol.* 148 (2), 1042–1054. doi: 10.1104/pp.108.126342

PubMed Abstract | CrossRef Full Text | Google Scholar

Nguyen, C. X., Paddock, K. J., Zhang, Z., Stacey, M. G. (2021). *GmKIX8-1* regulates organ size in soybean and is the causative gene for the major seed weight QTL *qSw17-1*. *New Phytol.* 229 (2), 920–934. doi: 10.1111/nph.16928

PubMed Abstract | CrossRef Full Text | Google Scholar

Ni, L., Liu, Y., Ma, X., Liu, T., Yang, X., Wang, Z., et al. (2023). Pan-3D genome analysis reveals structural and functional differentiation of soybean genomes. *Genome Biol.* 24 (1), 12. doi: 10.1186/s13059-023-02854-8

PubMed Abstract | CrossRef Full Text | Google Scholar

Ohyanagi, H., Sakata, K., Komatsu, S. (2012). Soybean proteome database 2012: update on the comprehensive data repository for soybean proteomics. *Front. Plant Sci.* 3. doi: 10.3389/fpls.2012.00110

CrossRef Full Text | Google Scholar

Pelletier, J. M., Kwong, R. W., Park, S., Le, B. H., Baden, R., Cagliari, A., et al. (2017). *LEC1* sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proc. Natl. Acad. Sci. U. S. A.* 114 (32), E6710–e6719. doi: 10.1073/pnas.1707957114

PubMed Abstract | CrossRef Full Text | Google Scholar

Ping, J., Liu, Y., Sun, L., Zhao, M., Li, Y., She, M., et al. (2014). *Dt2* is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean. *Plant Cell* 26 (7), 2831–2842. doi: 10.1105/tpc.114.126938

PubMed Abstract | CrossRef Full Text | Google Scholar

Qi, Z., Hou, M., Han, X., Lu, C., Jiang, H., Xin, D., et al. (2014). Identification of quantitative trait loci (QTLs) for seed protein concentration in soybean and analysis for additive effects and epistatic effects of QTLs under multiple environments. *Plant Breed.* 133 (4), 499–507. doi: 10.1111/pbr.12179

CrossRef Full Text | Google Scholar

Ray, D. K., Mueller, N. D., West, P. C., Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PloS One* 8 (6), e66428. doi: 10.1371/journal.pone.0066428

PubMed Abstract | CrossRef Full Text | Google Scholar

Reel, P. S., Reel, S., Pearson, E., Trucco, E., Jefferson, E. (2021). Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol. Adv.* 49, 107739. doi: 10.1016/j.biotechadv.2021.107739

PubMed Abstract | CrossRef Full Text | Google Scholar

Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M., Lepiniec, L. (2008). Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *Plant J.* 54 (4), 608–620. doi: 10.1111/j.1365-313X.2008.03461.x

PubMed Abstract | CrossRef Full Text | Google Scholar

Shen, B., Allen, W. B., Zheng, P., Li, C., Glassman, K., Ranch, J., et al. (2010). Expression of *ZmLEC1* and *ZmWRI1* increases seed oil production in maize. *Plant Physiol.* 153 (3), 980–987. doi: 10.1104/pp.110.157537

PubMed Abstract | CrossRef Full Text | Google Scholar

Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., et al. (2018a). *De novo* assembly of a Chinese soybean genome. *Sci. China Life Sci.* 61(8), 871–884. doi: 10.1007/s11427-018-9360-0

PubMed Abstract | CrossRef Full Text | Google Scholar

Shen, Y., Zhang, J., Liu, Y., Liu, S., Liu, Z., Duan, Z., et al. (2018b). DNA Methylation footprints during soybean domestication and improvement. *Genome Biol.* 19 (1), 128. doi: 10.1186/s13059-018-1516-z

PubMed Abstract | CrossRef Full Text | Google Scholar

Shen, Y., Zhou, G., Liang, C., Tian, Z. (2022). Omics-based interdisciplinarity is accelerating plant breeding. *Curr. Opin. Plant Biol.* 66, 102167. doi: 10.1016/j.pbi.2021.102167

PubMed Abstract | CrossRef Full Text | Google Scholar

Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., et al. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26 (3), 996–1008. doi: 10.1105/tpc.114.122739

PubMed Abstract | CrossRef Full Text | Google Scholar

Silva, E., Belinato, J. R., Porto, C., Nunes, E., Guimaraes, F., Meyer, M. C., et al. (2021). Soybean metabolomics based in mass spectrometry: decoding the plant's signaling and defense responses under biotic stress. *J. Agric. Food Chem.* 69 (26), 7257–7267. doi: 10.1021/acs.jafc.0c07758

PubMed Abstract | CrossRef Full Text | Google Scholar

Song, Q., Li, Q., Liu, Y., Zhang, F., Ma, B., Zhang, W., et al. (2013). Soybean *GmbZIP123* gene enhances lipid content in the seeds of transgenic *Arabidopsis* plants. *J. Exp. Bot.* 64 (14), 4329–4341. doi: 10.1093/jxb/ert238

PubMed Abstract | CrossRef Full Text | Google Scholar

Song, Y., Wang, X., Rose, R. J. (2017). Oil body biogenesis and biotechnology in legume seeds. *Plant Cell Rep.* 36 (10), 1519–1532. doi: 10.1007/s00299-017-2201-5

PubMed Abstract | CrossRef Full Text | Google Scholar

Tang, X., Su, T., Han, M., Wei, L., Wang, W., Yu, Z., et al. (2017). Suppression of extracellular invertase inhibitor gene expression improves seed weight in soybean (*Glycine max*). *J. Exp. Bot.* 68 (3), 469–482. doi: 10.1093/jxb/erw425

PubMed Abstract | CrossRef Full Text | Google Scholar

Tilman, D., Balzer, C., Hill, J., Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* 108 (50), 20260–20264. doi: 10.1073/pnas.1116437108

PubMed Abstract | CrossRef Full Text | Google Scholar

Wang, J., Chen, P., Wang, D., Shannon, G., Zeng, A., Orazaly, M., et al. (2015a). Identification and mapping of stable QTL for protein content in soybean seeds. *Mol. Breed.* 35 (3), 92. doi: 10.1007/s11032-015-0285-6

CrossRef Full Text | Google Scholar

Wang, X., Li, Y., Zhang, H., Sun, G., Zhang, W., Qiu, L. (2015b). Evolution and association analysis of *GmCYP78A10* gene with seed size/weight and pod number in soybean. *Mol. Biol. Rep.* 42 (2), 489–496. doi: 10.1007/s11033-014-3792-3

PubMed Abstract | CrossRef Full Text | Google Scholar

Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Yu, Y., et al. (2020b). Simultaneous changes in seed size, oil content and protein content driven by selection of *SWEET* homologues during soybean domestication. *Natl. Sci. Rev.* 7 (11), 1776–1786. doi: 10.1093/nsr/nwaa110

PubMed Abstract | CrossRef Full Text | Google Scholar

Wang, J., Schwab, R., Czech, B., Mica, E., Weigel, D. (2008). Dual effects of miR156-targeted *SPL* genes and *CYP78A5/KLUH* on plastochron length and organ size in *Arabidopsis* thaliana. *Plant Cell* 20 (5), 1231–1243. doi: 10.1105/tpc.108.058180

PubMed Abstract | CrossRef Full Text | Google Scholar

Wang, Q., Tang, J., Han, B., Huang, X. (2020a). Advances in genome-wide association studies of complex traits in rice. *Theor. Appl. Genet.* 133(5), 1415–1425. doi: 10.1007/s00122-019-03473-3

PubMed Abstract | CrossRef Full Text | Google Scholar

Wang, Z., Wang, Y., Shang, P., Yang, C., Yang, M., Huang, J., et al. (2022). Overexpression of soybean *GmWRI1a* stably increases the seed oil content in soybean. *Int. J. Mol. Sci.* 23 (9), 5084. doi: 10.3390/ijms23095084

PubMed Abstract | CrossRef Full Text | Google Scholar

Wang, S., Yokosho, K., Guo, R., Whelan, J., Ruan, Y., Ma, J., et al. (2019). The soybean sugar transporter *GmSWEET15* mediates sucrose export from endosperm to early embryo. *Plant Physiol.* 180 (4), 2133–2141. doi: 10.1104/pp.19.00641

PubMed Abstract | CrossRef Full Text | Google Scholar

Wang, H., Zhang, B., Hao, Y., Huang, J., Tian, A., Liao, Y., et al. (2007). The soybean dof-type transcription factor genes, *GmDof4* and *GmDof11*, enhance lipid content in the seeds of transgenic *Arabidopsis* plants. *Plant J.* 52 (4), 716–729. doi: 10.1111/j.1365-313X.2007.03268.x

PubMed Abstract | CrossRef Full Text | Google Scholar

Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., et al. (2015). QTL for seed protein and amino acids in the benning × danbaekkong soybean population. *Theor. Appl. Genet.* 128 (5), 839–850. doi: 10.1007/s00122-015-2474-4

PubMed Abstract | CrossRef Full Text | Google Scholar

Wei, Z., Pan, T., Zhao, Y., Su, B., Ren, Y., Qiu, L. (2020). The small GTPase Rab5a and its guanine nucleotide exchange factors are involved in post-golgi trafficking of storage proteins in developing soybean cotyledon. *J. Exp. Bot.* 71 (3), 808–822. doi: 10.1093/jxb/erz454

PubMed Abstract | CrossRef Full Text | Google Scholar

West, M., Yee, K. M., Danao, J., Zimmerman, J. L., Fischer, R. L., Goldberg, R. B., et al. (1994). LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. *Plant Cell* 6 (12), 1731–1745. doi: 10.1105/tpc.6.12.1731

PubMed Abstract | CrossRef Full Text | Google Scholar

Wilson, R. F. (2004). *"Seed composition," in Soybeans: Improvement, Production, and Uses, 3rd Edn, eds H. R. Boerma and J. E. Specht (Madison, WI: American Society of Agronomy)*, 621–677.

Google Scholar

Wilson, R. F. (2008). "Soybean: market driven research needs," in *Genetics and Genomics of Soybean, ed. G. Stacey* (New York, NY: Springer), 3–15. doi: 10.1007/978-0-387-72299-3_1

CrossRef Full Text | Google Scholar

Xu, C., Shanklin, J. (2016). Triacylglycerol metabolism, function, and accumulation in plant vegetative tissues. *Annu. Rev. Plant Biol.* 67, 179–206. doi: 10.1146/annurev-arplant-043015-111641

PubMed Abstract | CrossRef Full Text | Google Scholar

Yan, L., Hofmann, N., Li, S., Ferreira, M. E., Song, B., Jiang, G., et al. (2017). Identification of QTL with large effect on seed weight in a selective population of

soybean with genome-wide association and fixation index analyses. *BMC Genomics* 18 (1), 529. doi: 10.1186/s12864-017-3922-0

PubMed Abstract | CrossRef Full Text | Google Scholar

Yang, Y., Kong, Q., Lim, A. R. Q., Lu, S., Zhao, H., Guo, L., et al. (2022a). Transcriptional regulation of oil biosynthesis in seed plants: current understanding, applications, and perspectives. *Plant Commun.* 3 (5), 100328. doi: 10.1016/j.xplc.2022.100328

PubMed Abstract | CrossRef Full Text | Google Scholar

Yang, Y., La, T. C., Gillman, J. D., Lyu, Z., Joshi, T., Usovsky, M., et al. (2022b). Linkage analysis and residual heterozygotes derived near isogenic lines reveals a novel protein quantitative trait loci from a *Glycine soja* accession. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.938100

CrossRef Full Text | Google Scholar

Yang, H., Wang, W., He, Q., Xiang, S., Tian, D., Zhao, T., et al. (2019). Identifying a wild allele conferring small seed size, high protein content and low oil content using chromosome segment substitution lines in soybean. *Theor. Appl. Genet.* 132(10), 2793–2807. doi: 10.1007/s00122-019-03388-z

PubMed Abstract | CrossRef Full Text | Google Scholar

Yin, P., Ma, Q., Wang, H., Feng, D., Wang, X., Pei, Y., et al. (2020). *SMALL LEAF AND BUSHY1* controls organ size and lateral branching by modulating the stability of BIG SEEDS1 in *Medicago truncatula*. *New Phytol.* 226 (5), 1399–1412. doi: 10.1111/nph.16449

PubMed Abstract | CrossRef Full Text | Google Scholar

Yu, B., He, X., Tang, Y., Chen, Z., Zhou, L., Li, X., et al. (2023). Photoperiod controls plant seed size in a CONSTANS-dependent manner. *Nat. Plants* 9(2), 343–354. doi: 10.1038/s41477-023-01350-y

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, H., Goettel, W., Song, Q., Jiang, H., Hu, Z., Wang, M., et al. (2020). Selection of *GmSWEET39* for oil and protein improvement in soybean. *PloS Genet.* 16 (11), e1009114. doi: 10.1371/journal.pgen.1009114

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, M., Liu, S., Wang, Z., Yuan, Y., Zhang, Z., Liang, Q., et al. (2022). Progress in soybean functional genomics over the past decade. *Plant Biotechnol. J.* 20 (2), 256–282. doi: 10.1111/pbi.13682

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, J., Song, Q., Cregan, P., Jiang, G. (2016a). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (Glycine max). *Theor. Appl. Genet.* 129(1), 117–130. doi: 10.1007/s00122-015-2614-x

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, D., Sun, L., Li, S., Wang, W., Ding, Y., Swarm, S. A., et al. (2018a). Elevation of soybean seed oil content through selection for seed coat shininess. *Nat. Plants* 4 (1), 30–35. doi: 10.1038/s41477-017-0084-7

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, D., Wang, X., Li, S., Wang, C., Gosney, M. J., Mickelbart, M. V., et al. (2019a). A post-domestication mutation, *Dt2*, triggers systemic modification of divergent and convergent pathways modulating multiple agronomic traits in soybean. *Mol. Plant* 12 (10), 1366–1382. doi: 10.1016/j.molp.2019.05.010

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, J., Wang, X., Lu, Y., Bhusal, S. J., Song, Q., Cregan, P. B., et al. (2018b). Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. *Mol. Plant* 11 (3), 460–472. doi: 10.1016/j.molp.2017.12.016

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, T., Wu, T., Wang, L., Jiang, B., Zhen, C., Yuan, S., et al. (2019c). A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content. *Int. J. Mol. Sci.* 20 (23), 5915. doi: 10.3390/ijms20235915

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, W., Xu, W., Zhang, H., Liu, X., Cui, X., Li, S., et al. (2021). Comparative selective signature analysis and high-resolution GWAS reveal a new candidate gene controlling seed weight in soybean. *Theor. Appl. Genet.* 134 (5), 1329–1341. doi: 10.1007/s00122-021-03774-6

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, D., Zhang, H., Hu, Z., Chu, S., Yu, K., Lv, L., et al. (2019b). Artificial selection on *GmOLEO1* contributes to the increase in seed oil during soybean domestication. *PloS Genet.* 15 (7), e1008267. doi: 10.1371/journal.pgen.1008267

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, Y., Zhao, F., Li, Q., Niu, S., Wei, W., Zhang, W., et al. (2016b). Soybean *GmDREBL* increases lipid content in seeds of transgenic *Arabidopsis*. *Sci. Rep.* 6 (1), 34307. doi: 10.1038/srep34307

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhang, D., Zhao, M., Li, S., Sun, L., Wang, W., Cai, C., et al. (2017). Plasticity and innovation of regulatory mechanisms underlying seed oil content mediated by duplicated genes in the palaeopolyploid soybean. *Plant J.* 90 (6), 1120–1133. doi: 10.1111/tpj.13533

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhao, B., Dai, A., Wei, H., Yang, S., Wang, B., Jiang, N., et al. (2016). *Arabidopsis KLU* homologue *GmCYP78A72* regulates seed size in soybean. *Plant Mol. Biol.* 90 (1-2), 33–47. doi: 10.1007/s11103-015-0392-0

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhao, X., Dong, H., Chang, H., Zhao, J., Teng, W., Qiu, L., et al. (2019). Genome wide association mapping and candidate gene analysis for hundred seed weight in soybean [*Glycine max* (L.) Merrill]. *BMC Genomics* 20(1), 648. doi: 10.1186/s12864-019-6009-2

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33 (4), 408–414. doi: 10.1038/nbt.3096

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhu, W., Yang, C., Yong, B., Wang, Y., Li, B., Gu, Y., et al. (2022). An enhancing effect attributed to a nonsynonymous mutation in *SOYBEAN SEED SIZE 1*, a *SPINDLY*-like gene, is exploited in soybean domestication and improvement. *New Phytol.* 236 (4), 1375–1392. doi: 10.1111/nph.1846

PubMed Abstract | CrossRef Full Text | Google Scholar