

ĐÁNH GIÁ ĐA DẠNG DI TRUYỀN DỰA TRÊN KIỂU HÌNH QUẢ NGUỒN GEN CÀ CHUA BI QUA ẢNH

Nguyễn Trung Đức^{1*}, Nguyễn Thị Nguyệt Anh¹,
Phạm Quang Tuấn¹, Vũ Thị Thương², Nguyễn Việt Linh², Bùi Thị Thuý², Vũ Văn Liết²

¹*Viện Nghiên cứu và Phát triển cây trồng, Học viện Nông nghiệp Việt Nam*

²*Khoa Nông học, Học viện Nông nghiệp Việt Nam*

*Tác giả liên hệ: ntduc@vnua.edu.vn

Ngày nhận bài: 30.08.2022

Ngày chấp nhận đăng: 18.04.2023

TÓM TẮT

Mục tiêu của nghiên cứu này nhằm đánh giá đa dạng di truyền dựa trên kiểu hình quả của nguồn gen cà chua bi qua ảnh phục vụ chương trình chọn tạo giống cà chua ăn tươi chất lượng cao tại Việt Nam. Sáu mươi tám nguồn gen cà chua bi trồng trong vụ Đông 2021 tại Học viện Nông nghiệp Việt Nam được thu thập hình ảnh quả ở giai đoạn thu hoạch và trích xuất kiểu hình bằng phần mềm SmartGrain. So sánh độ chính xác với phương pháp đo đếm thủ công bằng mô hình hồi quy tuyến tính đơn giản. Kết quả cho thấy chiều dài và chiều rộng trích xuất qua ảnh có hệ số xác định cao (0,8136 và 0,7571, tương ứng) và sai số bình phương trung bình gốc thấp (2,947 và 2,345, tương ứng) so với phép đo thủ công. Các tính trạng quả trích xuất qua ảnh có tương quan thuận và chặt chẽ với khối lượng quả hơn so với các tính trạng đo đếm thủ công. Nguồn gen cà chua được phân thành 6 nhóm ở hệ số tương đồng 65,945 từ 5 tính trạng trích xuất qua ảnh. Phương pháp đánh giá kiểu hình quả qua ảnh có thể thay thế phương pháp đo đếm thủ công và ứng dụng trong các chương trình chọn tạo, khảo nghiệm giống cà chua và trên các giống cây trồng khác.

Từ khóa: Cà chua bi, nguồn gen, phân tích ảnh, phân tích thành phần chính, phân tích cụm.

Assessment of Genetic Diversity on Fruit Traits of Cherry Tomato Germplasm Based on Images

ABSTRACT

This study aimed to assess the fruit trait diversity of cherry tomato germplasm through image-based phenotyping for the nutritious cherry tomato breeding program in Vietnam. Fruit images of sixty-eight cherry tomato accessions grown in winter crop season 2021 at the Vietnam National University of Agriculture were collected at the harvest stage and then were analyzed by SmartGrain software. Comparing the accuracy with the manual measurement using a simple linear regression model, the obtained results show that the fruit length and width extracted from the image had a high determination coefficient (0.81 and 0.76, respectively) and low root mean square error (2.95 and 2.35, respectively) compared with manual measurements. The image-extracted fruit traits were more closely and positively correlated with fruit weight as compared to manual measurements. The tomato germplasm were classified into 6 groups at the similarity coefficient of 65.95 from 5 image-extracted fruit traits. Fruit image-based phenotyping can replace the manual measurement and apply in tomato breeding, testing system, and further apply to other cultivars.

Keywords: Cherry tomato, germplasm, image analysis, principal component analysis, cluster analysis.

1. ĐẶT VẤN ĐỀ

Cà chua bi (*Solanum lycopersicum* var. *Cerasiforme*) được cho là tổ tiên của cà chua hiện đại ngày nay (Ranc & cs., 2008). Đây là loại cà chua phổ biến vì kích thước quả nhỏ,

hương vị tinh tế, chất lượng quả cao và ngày càng được ưa chuộng như một loại rau ăn tươi (Liu & cs., 2018). Mặc dù có thể trồng được trên đồng ruộng nhưng với năng suất, chất lượng và giá trị kinh tế cao, các giống cà chua bi phần lớn được canh tác theo hướng nông

nghiệp công nghệ cao trong nhà kính, nhà màng (He & cs., 2021; Kong & cs., 2021). Các giống cà chua bi trồng phổ biến trong nước như Thúy Hồng (quả đỏ), Nova (quả vàng), Chocolate (quả tím), Indigo Rose (tím, đen) đều là giống lai và hoàn toàn nhập nội với giá thành hạt giống cao (khoảng 2.000-4.500 đồng/hạt). Chọn tạo giống cà chua tại Việt Nam đã phát triển được giống cà chua bi đỏ HT144 (Nguyễn Hồng Minh & cs., 2011), cà chua quả to lai VT15 (Đoàn Xuân Cảnh & cs., 2021), giống cà chua thuần TP130 và TP135 kháng bệnh xoắn vàng lá (Tống Văn Hải & cs., 2021). Trong các nghiên cứu này, kiểu hình quả đóng vai trò quan trọng và đều được đo đếm thủ công. Đến nay, phương pháp đo đếm này vẫn hiệu quả nhưng tốn thời gian, công sức và thường có sai số lớn. Với sự tiến bộ của khoa học máy tính cùng các công cụ phân tích hình ảnh miễn phí, các nghiên cứu đánh giá kiểu hình quả, bắp, hạt hoàn toàn có thể ứng dụng ngay với phương pháp đơn giản, chi phí thấp và hiệu quả cao (Nguyễn Trung Đức & cs., 2022).

Thông tin kiểu hình quả không thể thiếu trong nghiên cứu đa dạng di truyền (Nankar & cs., 2020); xác định tính khác biệt, đồng nhất, ổn định phục vụ bảo hộ và công nhận lưu hành các giống cà chua mới; ảnh hưởng tới thị hiếu người tiêu dùng và giá trị thị trường (Causse & cs., 2010). Vì vậy, nhà chọn giống luôn quan tâm tới phương pháp xác định hình dạng, kích thước quả một cách chính xác, khách quan, với số lượng mẫu lớn trong thời gian ngắn và có thể tái lập. Trên thế giới đã có một số phần mềm phân tích tính trạng quả, hạt với số lượng mẫu lớn với độ chính xác cao qua ảnh RGB được phát triển như Tomato Analyzer (Rodriguez & cs., 2010), SmartGrain (Tanabata & cs., 2012), kết hợp thư viện thị giác máy tính OpenCV và phương pháp học sâu (Zhu & cs., 2022). Trong các phần mềm này, hiệu quả, đơn giản và dễ sử dụng nhất là phần mềm SmartGrain được phát triển bởi Tanabata & cs. (2012) với thuật toán tự động xác định các đường viền của hạt, quả dựa trên thông tin hình ảnh (mục tiêu cần phân tích và ảnh nền) và xuất ra giá trị của các tính trạng được phần mềm đo lường như diện tích

quả (AS), chu vi quả (P), chiều dài quả (L), chiều rộng quả (W) và tỷ lệ chiều dài trên chiều rộng (LWR) dưới định dạng Excel.

Nguồn gen cà chua bi được lưu trữ, phát triển tại Viện Nghiên cứu và Phát triển cây trồng, Học viện Nông nghiệp Việt Nam trong hơn một thập kỷ qua khá đa dạng và số lượng không ngừng tăng khiến cho công tác đo đếm các tính trạng số lượng tốn nhiều chi phí, thời gian và công sức. Do vậy, nghiên cứu này áp dụng phương pháp đánh giá kiểu hình quả qua ảnh nhằm xác nhận sự chính xác của phương pháp mới, qua đó sử dụng để tìm hiểu thông tin về sự đa dạng di truyền hình dạng quả trên nguồn gen cà chua bi tại Việt Nam.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu

Vật liệu nghiên cứu gồm 68 nguồn vật liệu cà chua bi ký hiệu từ CTM01 đến CTM68 có nguồn gốc trong nước và nhập nội được lưu trữ, duy trì và phát triển tại Viện Nghiên cứu và Phát triển Cây trồng, Học viện Nông nghiệp Việt Nam (Hình 1).

2.2. Bố trí thí nghiệm

Thí nghiệm đồng ruộng được bố trí tuần tự, không nhắc lại, diện tích ô thí nghiệm 15m² trong vụ Đông 2021 tại Viện Nghiên cứu và Phát triển cây trồng, Học viện Nông nghiệp Việt Nam. Các vật liệu cà chua được gieo bầu, trồng sau khi cây được 1 tháng tuổi, chăm sóc, bón phân, phòng trừ sâu bệnh hại theo quy chuẩn QCVN 01-63:2011/BNNT.

2.3. Đo đếm các tính trạng thủ công

Mỗi nguồn gen cà chua chọn 13-15 quả ở lứa thứ 2-3 khi quả chín hoàn toàn để đo đếm thủ công và tính giá trị trung bình để làm số liệu so sánh với kết quả phân tích từ các phần mềm trích xuất hình ảnh, bao gồm 4 tính trạng: chiều dài quả (mm), chiều rộng quả (mm), chỉ số dạng quả là tỉ lệ chiều dài/chiều rộng, và khối lượng quả trung bình (g).



Hình 1. Sáu mươi tám nguồn gen cà chua bi nghiên cứu

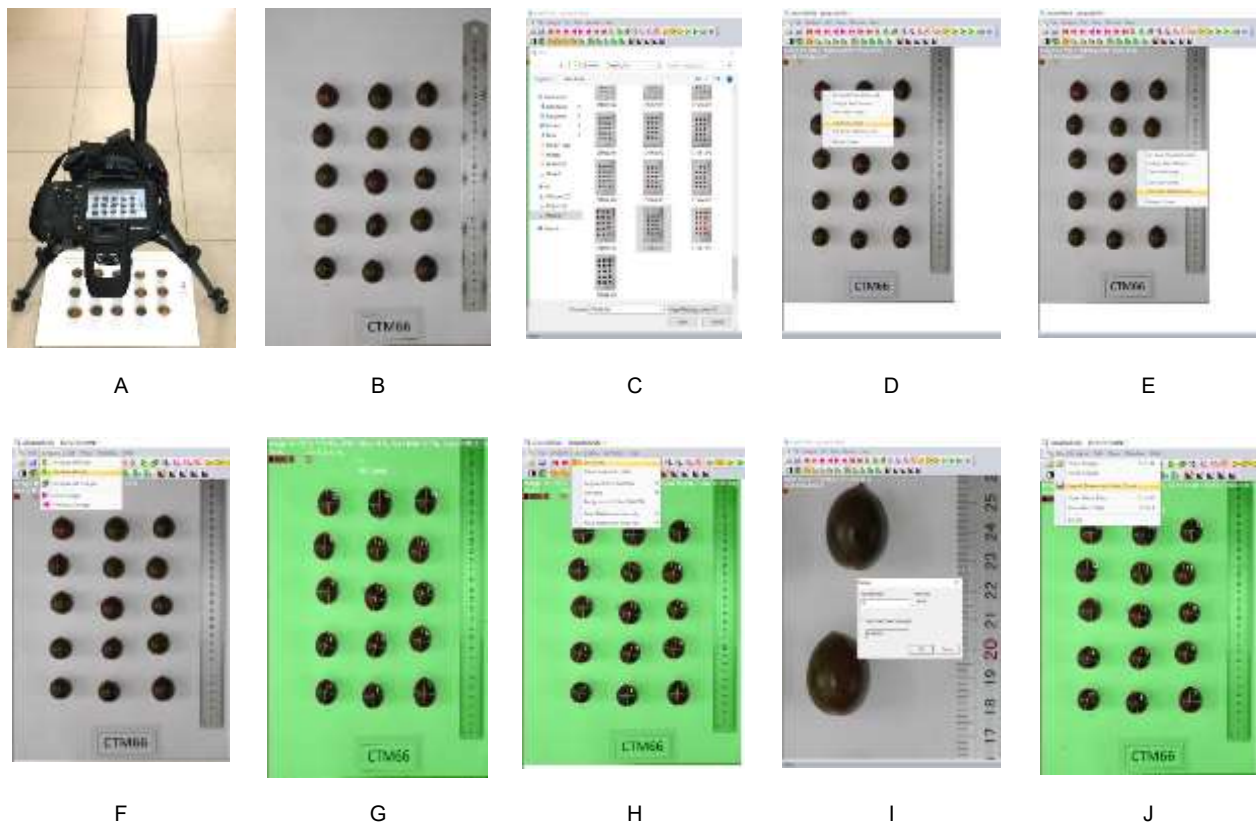
Bảng 1. Đặc điểm các tính trạng quả đo đếm thủ công của nguồn gen cà chua bi

Dòng	Chiều dài (D) (mm)		Chiều rộng (R) (mm)		Chỉ số dạng quả (D/R)		Khối lượng TB quả (g)	Dòng	Chiều dài (D) (mm)		Chiều rộng (R) (mm)		Chỉ số dạng quả (D/R)		Khối lượng TB quả (g)
	TB	SE	TB	SE	TB	SE			TB	SE	TB	SE	TB	SE	
CTM01	28,69	0,54	25,68	1,42	1,12	0,04	11,67	CTM35	30,70	2,40	24,87	3,04	1,23	0,03	12,00
CTM02	34,70	1,18	27,77	2,63	1,25	0,06	13,00	CTM36	29,41	0,52	28,80	1,06	1,02	0,03	13,00
CTM03	28,21	1,50	26,96	0,46	1,05	0,04	12,33	CTM37	35,82	2,81	27,65	3,53	1,30	0,02	12,00
CTM04	35,55	1,55	33,32	1,78	1,07	0,04	13,33	CTM38	36,03	1,94	28,60	1,95	1,26	0,02	13,67
CTM05	24,77	4,82	21,16	3,37	1,17	0,11	6,67	CTM39	29,95	0,56	24,65	0,89	1,22	0,04	9,33
CTM06	28,31	2,22	21,45	2,81	1,32	0,04	8,00	CTM40	31,49	0,70	25,90	0,75	1,22	0,01	12,00

Đánh giá đa dạng di truyền dựa trên kiểu hình quả nguồn gen cà chua bi qua ảnh

CTM07	34,02	5,59	31,07	1,65	1,09	0,16	12,00	CTM41	28,90	2,25	23,15	2,19	1,25	0,06	8,00
CTM08	37,68	1,16	28,88	5,71	1,30	0,15	16,00	CTM42	32,78	2,22	28,75	1,24	1,14	0,05	13,33
CTM09	29,66	2,51	24,51	3,60	1,21	0,03	9,67	CTM43	30,07	0,96	24,99	0,55	1,20	0,06	9,67
CTM10	27,68	3,33	25,66	3,44	1,08	0,07	12,00	CTM44	35,39	0,70	29,53	2,81	1,20	0,07	14,67
CTM11	34,21	0,89	29,32	1,55	1,17	0,08	16,00	CTM45	32,46	5,07	25,47	7,10	1,27	0,07	9,67
CTM12	35,59	7,02	32,00	5,34	1,11	0,12	15,00	CTM46	25,03	0,86	22,99	1,71	1,09	0,03	8,33
CTM13	33,97	2,70	29,44	4,16	1,15	0,04	13,67	CTM47	27,96	2,25	27,08	2,35	1,03	0,05	12,00
CTM14	34,71	5,82	25,98	7,07	1,34	0,04	12,33	CTM48	30,08	1,46	27,53	0,70	1,09	0,03	11,33
CTM15	50,53	1,44	40,08	3,97	1,26	0,06	48,00	CTM49	27,60	5,26	22,35	1,27	1,23	0,33	7,33
CTM16	35,35	2,55	31,41	1,68	1,13	0,04	20,00	CTM50	26,85	2,56	24,35	3,40	1,10	0,03	7,67
CTM17	47,26	0,84	29,22	1,86	1,62	0,11	26,00	CTM51	26,71	2,63	25,51	3,59	1,05	0,03	8,33
CTM18	34,90	3,68	29,84	5,19	1,17	0,22	12,67	CTM52	28,78	2,49	24,27	2,79	1,19	0,04	9,33
CTM19	41,97	1,55	42,13	9,90	1,00	0,22	38,67	CTM53	22,85	4,29	17,89	4,34	1,28	0,08	6,33
CTM20	33,78	5,20	32,75	2,35	1,03	0,11	16,00	CTM54	32,78	4,77	29,06	4,70	1,13	0,03	13,67
CTM21	56,31	6,21	34,06	9,51	1,65	0,05	28,00	CTM55	32,02	2,93	31,12	2,76	1,03	0,03	14,33
CTM22	54,65	4,44	43,39	1,25	1,26	0,11	34,67	CTM56	26,02	5,40	23,69	4,61	1,10	0,15	8,00
CTM23	52,56	5,88	37,18	1,13	1,41	0,21	34,67	CTM57	25,65	1,85	17,29	0,21	1,48	0,16	9,33
CTM24	29,10	0,34	26,05	1,70	1,12	0,05	12,67	CTM58	28,25	4,88	27,75	3,76	1,02	0,09	12,00
CTM25	32,05	1,04	28,64	1,06	1,12	0,07	13,00	CTM59	36,44	1,31	29,33	3,34	1,24	0,06	14,00
CTM26	25,85	1,24	21,15	1,26	1,22	6,85	12,67	CTM60	31,29	2,95	28,88	1,30	1,08	0,07	12,67
CTM27	33,35	1,72	30,65	1,76	1,09	0,01	14,00	CTM61	34,58	2,62	30,95	2,62	1,12	0,10	13,33
CTM28	28,44	2,18	25,93	2,72	1,10	0,03	5,67	CTM62	36,65	3,71	29,07	3,21	1,26	0,11	14,00
CTM29	26,99	3,05	24,39	3,78	1,11	0,03	10,00	CTM63	29,99	1,71	26,07	2,73	1,15	0,03	10,00
CTM30	44,72	1,65	35,79	1,42	1,25	0,10	19,00	CTM64	34,55	1,31	27,05	1,63	1,28	0,12	12,00
CTM31	31,71	5,00	27,86	4,89	1,14	0,06	12,00	CTM65	35,25	1,56	31,88	2,62	1,11	0,07	17,33
CTM32	26,38	0,93	23,24	1,36	1,14	0,01	12,00	CTM66	26,76	0,40	22,37	1,80	1,20	0,06	8,00
CTM33	37,55	1,05	29,24	3,53	1,28	0,09	13,33	CTM67	25,93	5,16	23,47	4,05	1,10	0,07	7,95
CTM34	37,74	1,63	29,28	2,11	1,29	0,07	15,33	CTM68	37,35	4,52	42,19	3,56	0,89	0,05	28,95

Ghi chú: TB: Trung bình, SE: Sai số chuẩn.



Hình 2. Các bước trích xuất kiểu hình quả qua ảnh

2.4. Thu thập, phân tích tính trạng quả qua hình ảnh

Mẫu quả của các dòng cà chua ở giai đoạn thu hoạch (13-15 quả/nguồn gen) được chụp bằng máy ảnh Canon EOS 1500 với nền trắng. Máy ảnh được giữ cố định với tham số F-stop $f/5.6$, thời gian phơi sáng $1/200$ giây, ISO-3200, tiêu cự 33cm và chuẩn màu sRGB. Mỗi hình ảnh RGB có kích thước 4.000×6.000 pixel (Hình 2A, B). Phân tích và trích xuất kiểu hình qua hình ảnh quả bằng phần mềm SmartGrain (Tanabata & cs., 2012).

Các bước chính gồm: mở hình ảnh cần phân tích (Hình 2C); chọn màu sắc quả (Hình 2D); chọn màu sắc phong nền (Hình 1E); tiếp đó chọn vùng ảnh cần phân tích (Hình 1F), thu được kết quả (Hình 1G); tiếp đó chọn thước cho hình ảnh dựa vào thước trên hình (H, I) và cuối cùng là xuất file “.csv” lưu kết quả phân tích (Hình 1J). Các tính trạng trích xuất qua hình ảnh bằng

phần mềm SmartGrain gồm: diện tích quả (AS, mm^2), chu vi quả (PL, mm), chiều dài quả (L, mm), chiều rộng quả (W, mm) và chỉ số dạng quả (LWR).

2.5. Phân tích số liệu

Số liệu kiểu hình tổng hợp trên phần mềm Microsoft Excel. Phân tích tương quan sử dụng gói “*metan*” trên phần mềm R 4.1.1 (R Development Core Team, 2022). Biểu đồ phân nhóm đa dạng di truyền quả được xây dựng dựa trên khoảng cách Euclidean trên phần mềm Minitab 19.0.

So sánh sự khác biệt giữa tính trạng chiều dài quả, chiều rộng quả, chỉ số dạng quả trích xuất qua ảnh với các phép đo thủ công bằng mô hình hồi quy tuyến tính đơn giản, sử dụng tính trạng trích xuất qua ảnh ở trục y và phép đo thủ công tương ứng ở trục x. Hệ số xác định (R^2), sai số bình phương trung bình gốc (RMSE) được tính bằng các phương trình sau:

$$R^2 = 1 - \frac{\sum_i (x_i - y_i)^2}{\sum_i (x_i - \bar{y})^2}$$

$$RMSE = \sqrt{\frac{\sum_i (x_i - y_i)^2}{n}}$$

Trong đó n là tổng số phép đo; x_i là kết quả đo thủ công thứ i; y_i là kết quả trích xuất từ phần mềm SmartGrain thứ i, tương ứng và $\bar{y}$ là giá trị trung bình của các phép đo trích xuất từ phần mềm SmartGrain.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Đặc điểm kiểu hình quả đo đếm thủ công và trích xuất qua ảnh

Kết quả thống kê đặc điểm kiểu hình quả đo đếm thủ công (Bảng 1 và 3) cho thấy các dòng cà chua bi có sự sai khác và biến động đáng kể về chỉ tiêu chiều dài quả đo đếm thủ công, dao động từ 22,85mm (CTM19) đến 56,31 mm (CTM21), trung bình $33,11 \pm 0,84$ mm, với hệ số biến động 18,30. Chiều rộng quả đo đếm thủ công của các nguồn gen chua bi dao động từ 17,29mm (CTM57) đến 43,39mm (CTM22), trung bình $28,09 \pm 0,62$ mm, với hệ số biến động 21,02. Chỉ số dạng quả tính từ phép đo thủ công dao động từ 0,89 (CTM68) đến 1,65 (CTM21), trung bình đạt $1,18 \pm 0,02$. Khối lượng trung bình quả đo đếm thủ công của các dòng cà chua bi dao động từ 5,67g (CTM28) đến 48,00g (CTM15), trung bình đạt 14,17g.

Kết quả đánh giá kiểu hình về đặc kiểu hình trích xuất qua ảnh cho thấy các nguồn gen cà chua bi có diện tích quả dao động từ 414,32mm² (CTM57) đến 2097,02mm² (CTM15), trung bình đạt 741,98mm² với hệ số biến động 40,54. Chu vi quả dao động từ 80,39mm (CTM53) đến 213,97mm (CTM68), trung bình đạt 108,31mm với hệ số biến động 22,21. Kết quả thống kê mô tả cũng chỉ ra không có sự sai khác lớn giữa chiều dài, chiều rộng và chỉ số dạng quả trích xuất qua ảnh với phép đo đếm thủ công (Bảng 3).

Việc định lượng chính xác các đặc điểm hình thái quả, hạt có thể thúc đẩy sự tiến bộ của

các lĩnh vực nghiên cứu thực vật bao gồm di truyền học, sinh lý học, phân tích chức năng gen và ứng dụng trong chọn giống cây trồng (Tanabata & cs., 2012; Zhu & cs., 2022). Sự thành công của một chương trình chọn giống phụ thuộc vào các biến dị có trong quần thể. Kết quả phân tích kiểu hình quả ở nghiên cứu này cho thấy khoảng dao động và hệ số biến động lớn, trong đó lớn nhất ở các tính trạng khối lượng quả (54,55%) và diện tích quả (40,54%) biểu thị mức độ đa dạng cao về kích thước quả trên 68 nguồn gen cà chua. Kết quả của nghiên cứu cũng cung cấp thông tin quan trọng cho các chương trình tạo giống cà chua bi tại Học viện Nông nghiệp Việt Nam. Tuy nhiên, để đạt được kết luận chắc chắn về tính đa dạng di truyền, cần tiến hành thêm các nghiên cứu đánh giá đa dạng di truyền ứng dụng chỉ thị phân tử. Mặt khác, so sánh độ chính xác giữa phép đo trích xuất qua ảnh và phép đo thủ công sẽ là cơ sở để xuất áp dụng phương pháp trích xuất kiểu hình qua ảnh trong các nghiên cứu tiếp theo.

3.2. So sánh phép đo trích xuất qua ảnh và phép đo thủ công

So sánh sự chính xác giữa kết quả chiều dài, chiều rộng quả trích xuất qua ảnh và chiều dài và chiều rộng quả đo đếm thủ công sử dụng mô hình hồi quy tuyến tính đơn giản cho thấy cả hai mô hình đều có hệ số xác định R^2 cao ($R^2 = 0,8136$; $R^2 = 0,7571$; tương ứng) và sai số bình phương trung bình gốc RMSE thấp (RMSE = 2,947 và 2,345; tương ứng) (Hình 3). Nghiên cứu của Zhu & cs. (2022) đã sử dụng tham số RMSE để xác định sự chính xác giữa phương pháp phân tích học sâu từ hình ảnh quả cà chua với kết quả đo đếm thủ công. Như vậy, phép đo trích xuất qua ảnh có thể dùng để thay thế các phép đo chiều dài và chiều rộng thủ công. Phân tích tương quan tiếp tục làm sáng tỏ mối quan hệ giữa các tính trạng trích xuất qua ảnh, đo đếm thủ công với khối lượng quả.

Kết quả phân tích tương quan đã chỉ ra các tính trạng quả trích xuất qua ảnh có tương quan thuận và chặt với khối lượng quả hơn so với các tính trạng hình dạng và kích thước quả đo đếm thủ công (Hình 2).

Bảng 2. Đặc điểm tính trạng quả vật liệu cà chua bi trích xuất qua ảnh

Dòng	Diện tích (mm ²)		Chu vi (mm)		Chiều dài (D) (mm)		Chiều rộng (R) (mm)		Chỉ số dạng quả (D/R)	
	TB	SE	TB	SE	TB	SE	TB	SE	TB	SE
CTM01	568,60	77,99	94,68	6,59	29,16	2,19	25,46	1,75	1,15	0,05
CTM02	698,76	81,49	106,36	6,05	33,88	3,02	27,44	1,21	1,23	0,09
CTM03	538,37	57,86	101,74	10,37	27,84	1,33	26,23	1,53	1,06	0,05
CTM04	716,98	208,11	103,87	16,92	31,86	5,66	28,56	4,05	1,11	0,09
CTM05	461,35	93,80	82,38	8,98	26,25	2,76	22,38	2,77	1,18	0,09
CTM06	438,49	123,73	93,47	17,97	27,59	4,10	20,66	3,29	1,34	0,10
CTM07	583,19	107,50	97,87	9,95	31,60	2,53	25,35	3,11	1,26	0,09
CTM08	866,50	134,48	121,06	10,22	38,18	3,58	29,07	1,90	1,31	0,09
CTM09	532,38	151,35	92,63	23,53	28,61	6,89	24,06	5,71	1,19	0,11
CTM10	619,49	83,80	100,77	7,85	30,25	2,21	26,93	2,23	1,13	0,06
CTM11	688,25	121,54	112,95	10,83	32,82	3,13	28,79	2,48	1,14	0,05
CTM12	777,95	132,67	111,88	10,49	36,13	3,66	28,05	2,50	1,29	0,08
CTM13	693,90	171,50	105,23	13,63	31,76	3,80	28,99	4,09	1,10	0,07
CTM14	901,63	161,57	122,65	13,86	38,64	4,28	31,31	3,50	1,24	0,11
CTM15	2097,02	390,83	194,68	32,67	59,58	6,10	46,41	5,56	1,29	0,10
CTM16	1003,09	124,06	123,66	8,49	37,79	2,75	34,55	1,95	1,09	0,05
CTM17	1058,61	158,28	138,04	14,66	46,55	4,35	29,98	3,00	1,56	0,16
CTM18	664,53	202,88	101,92	27,44	31,45	8,43	26,74	7,21	1,18	0,07
CTM19	1.552,48	234,51	158,63	12,26	47,63	4,15	42,81	3,36	1,11	0,06
CTM20	702,01	156,14	107,06	12,10	31,71	3,47	29,24	3,47	1,09	0,06
CTM21	1.360,08	334,42	151,89	25,28	50,10	8,93	34,76	3,50	1,44	0,20
CTM22	1.468,65	509,34	156,03	28,49	48,08	8,84	39,28	5,28	1,22	0,14
CTM23	1.443,44	297,10	159,52	26,15	51,30	7,84	37,03	3,05	1,38	0,17
CTM24	605,05	77,39	102,08	12,74	30,48	2,37	26,79	1,78	1,14	0,05
CTM25	670,05	99,30	103,16	8,90	31,44	2,77	28,28	2,22	1,11	0,07
CTM26	635,29	91,71	98,50	7,65	30,74	2,18	26,99	2,38	1,14	0,05
CTM27	718,80	177,40	107,62	16,26	32,07	3,90	29,26	3,60	1,10	0,05
CTM28	581,66	144,77	96,87	15,74	29,48	3,47	25,88	3,32	1,14	0,07
CTM29	529,21	110,04	94,17	9,90	27,89	2,59	25,12	2,74	1,11	0,04
CTM30	922,23	141,35	122,78	9,54	40,12	3,27	31,01	2,60	1,30	0,10
CTM31	641,61	176,92	101,58	17,04	31,08	4,09	27,14	4,22	1,15	0,05
CTM32	625,42	90,27	99,62	8,48	29,86	2,48	27,89	2,02	1,07	0,04
CTM33	732,21	159,74	108,46	13,28	35,27	3,86	27,21	3,72	1,30	0,09
CTM34	807,58	134,38	112,06	9,19	34,38	3,05	30,69	2,56	1,12	0,06
CTM35	616,92	82,72	98,83	7,74	31,07	2,00	26,52	2,48	1,18	0,07
CTM36	684,59	83,65	103,64	7,92	31,02	1,81	29,08	2,16	1,07	0,03
CTM37	639,96	91,76	102,88	7,63	33,68	2,74	26,42	1,88	1,28	0,06
CTM38	763,51	153,73	109,93	11,85	34,08	3,69	30,15	3,38	1,13	0,07
CTM39	577,15	71,74	94,76	6,33	29,17	1,99	26,91	1,83	1,08	0,04
CTM40	655,84	82,38	100,42	7,44	31,39	2,27	27,92	1,86	1,13	0,05
CTM41	488,00	74,42	87,29	6,81	27,36	1,77	23,78	2,15	1,15	0,06
CTM42	723,30	125,22	104,97	9,76	32,39	2,75	29,52	2,97	1,10	0,04

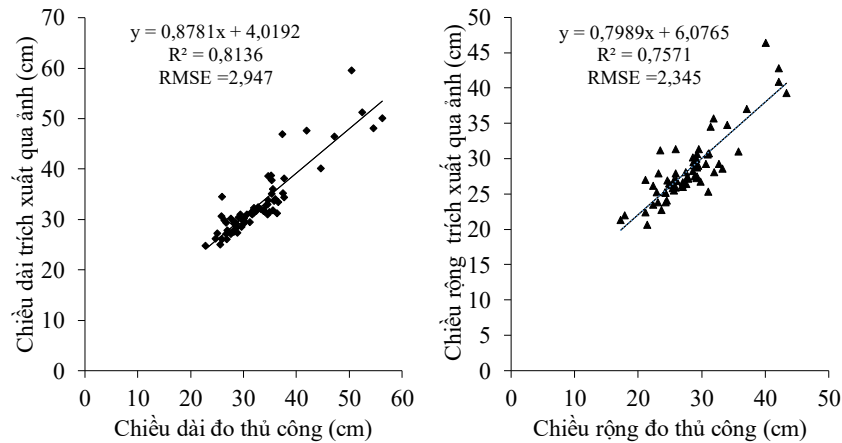
Đánh giá đa dạng di truyền dựa trên kiểu hình quả nguồn gen cà chua bi qua ảnh

Dòng	Diện tích (mm ²)		Chu vi (mm)		Chiều dài (D) (mm)		Chiều rộng (R) (mm)		Chỉ số dạng quả (D/R)	
	TB	SE	TB	SE	TB	SE	TB	SE	TB	SE
CTM43	596,03	101,85	97,15	8,41	30,23	2,51	26,21	2,38	1,16	0,05
CTM44	821,89	80,14	113,88	6,35	35,17	2,29	31,35	1,54	1,12	0,05
CTM45	627,18	111,15	98,96	9,97	31,73	3,18	26,40	2,63	1,21	0,09
CTM46	524,92	84,16	89,52	7,94	27,24	2,22	25,27	1,98	1,08	0,04
CTM47	574,46	178,95	90,93	23,59	27,58	7,03	25,99	6,73	1,07	0,06
CTM48	641,77	144,34	97,70	11,78	29,69	3,34	28,04	3,31	1,06	0,03
CTM49	495,29	105,67	86,46	10,49	27,12	2,76	23,50	3,55	1,17	0,11
CTM50	481,21	93,98	85,69	8,59	26,11	2,28	23,81	2,70	1,10	0,05
CTM51	536,81	87,54	90,02	7,69	27,19	2,18	25,81	2,30	1,05	0,03
CTM52	535,88	67,94	89,73	5,95	28,51	1,54	25,27	2,06	1,13	0,07
CTM53	420,39	71,11	80,39	6,90	24,83	2,18	21,99	1,94	1,13	0,06
CTM54	741,73	188,77	106,02	13,70	32,45	3,92	29,69	4,27	1,10	0,06
CTM55	761,10	183,53	107,53	13,96	32,28	3,77	30,79	3,78	1,05	0,03
CTM56	495,48	189,57	84,34	25,09	26,14	7,81	22,70	6,45	1,16	0,09
CTM57	414,32	84,16	80,42	9,86	25,02	2,79	21,33	2,18	1,17	0,06
CTM58	616,96	177,15	97,12	14,31	29,32	4,08	27,22	4,05	1,08	0,05
CTM59	714,02	289,50	103,17	38,63	31,30	11,51	27,53	10,19	1,14	0,06
CTM60	655,66	209,29	99,26	26,60	29,52	7,85	27,49	7,40	1,09	0,07
CTM61	767,15	115,41	112,00	10,03	33,03	2,51	30,55	2,33	1,08	0,04
CTM62	761,85	149,50	111,50	12,36	33,52	3,27	30,04	3,16	1,12	0,04
CTM63	606,81	83,01	97,83	6,63	29,75	1,88	27,03	2,09	1,10	0,05
CTM64	638,38	195,25	99,37	27,01	31,04	8,41	26,67	7,38	1,18	0,10
CTM65	1064,02	166,72	131,58	10,88	38,79	3,04	35,69	3,18	1,09	0,03
CTM66	611,48	166,23	96,32	24,76	29,43	7,59	26,12	6,63	1,12	0,06
CTM67	830,92	99,40	115,41	6,56	34,55	2,26	31,18	1,78	1,11	0,03
CTM68	1465,06	225,81	213,97	43,70	46,99	4,06	40,85	3,13	1,15	0,06

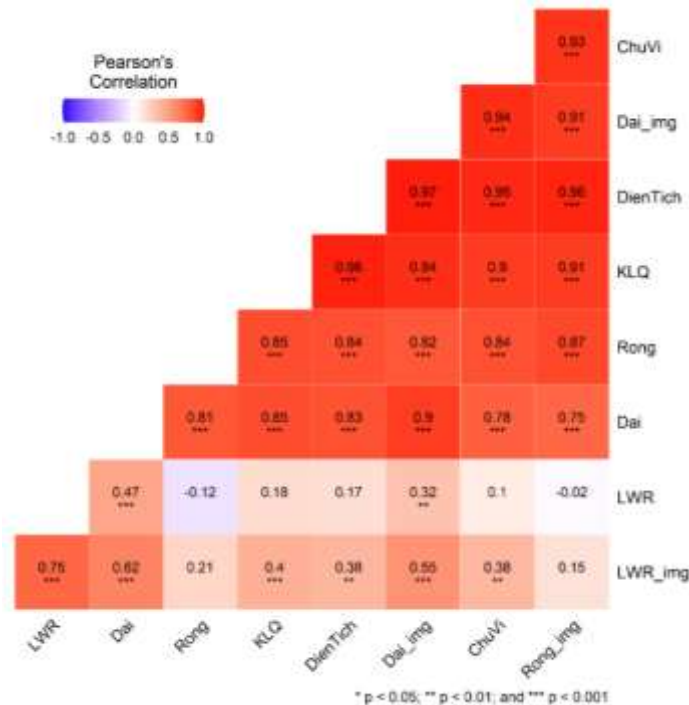
Ghi chú: TB: Trung bình, SE: Sai số chuẩn.

Bảng 3. Thống kê mô tả các tính trạng đo đếm bằng phương pháp truyền thống và trích xuất qua ảnh

Tính trạng	Giá trị cao nhất	Giá trị thấp nhất	Giá trị trung bình	Hệ số biến động (%)	Độ lệch chuẩn	Sai số chuẩn
<i>Các tính trạng đo đếm truyền thống</i>						
Chiều dài quả (D) (mm)	56,31	22,85	33,11	18,30	1,73	0,84
Chiều rộng quả (R) (mm)	43,39	17,29	28,09	21,02	5,14	0,62
Chỉ số dạng quả (D/R)	1,65	0,89	1,18	11,02	0,13	0,02
Khối lượng quả (g)	48,00	5,67	14,17	54,55	7,73	0,94
<i>Các tính trạng trích xuất qua ảnh</i>						
Diện tích quả (mm ²)	2097,02	414,32	741,98	40,54	300,77	36,47
Chu vi quả (mm)	213,97	80,39	108,31	22,21	24,05	2,92
Chiều dài quả (D) (mm)	59,58	24,83	33,09	20,47	6,77	0,82
Chiều rộng quả (R) (mm)	46,41	20,66	28,52	16,56	4,72	0,57
Chỉ số dạng quả (D/R)	1,56	1,05	1,16	8,29	0,10	0,01



Hình 3. So sánh phép đo trích xuất qua ảnh và phép đo thủ công



Ghi chú: Dai: Chiều dài quả đo đếm thủ công; Rong: Chiều rộng quả đo đếm thủ công; KLQ: Khối lượng quả đo đếm thủ công; DienTich: Diện tích trích xuất qua ảnh; Dai_img: Chiều dài trích xuất qua ảnh; Rong_img: Chiều rộng trích xuất qua ảnh; LWR_img: Chỉ số dạng quả xuất trích xuất qua ảnh.

Hình 4. Biểu đồ nhiệt biểu thị tương quan của các tính trạng nghiên cứu

Kết quả phân tích thành phần chính trên các tính trạng theo dõi cho thấy tổng hai thành phần chính đầu tiên đóng góp 92,5%, chín tính trạng đều có đóng góp trên 10% cho thấy sự quan trọng của các tính trạng quả (Hình 5). Các tính trạng kiểu hình trích xuất qua ảnh có chiều dài vector dài hơn cho thấy đóng góp nhiều hơn so với các tính trạng đo đếm bằng phương pháp

đo đếm thủ công. Sự tương quan thuận và rất chặt giữa khối lượng trung bình quả và tính trạng kiểu hình trích xuất qua ảnh (Hình 4) cùng với sự đóng góp nhiều hơn so với các tính trạng đo đếm thủ công (Hình 5) là cơ sở áp dụng các phương pháp học máy dự đoán khối lượng quả từ tập dữ liệu kiểu hình trích xuất qua ảnh trong nghiên cứu tiếp theo. Trên thế giới,

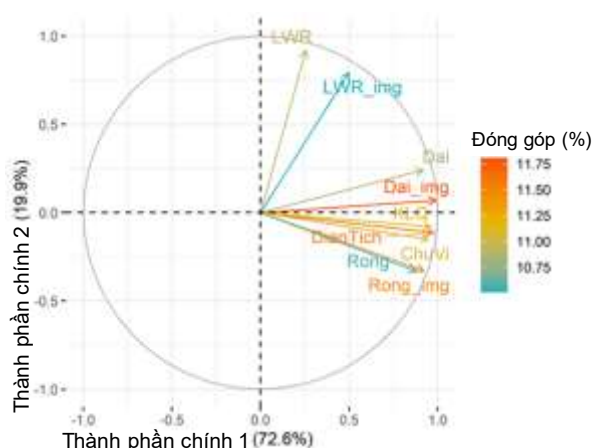
phương pháp phân tích thành phần chính đã được sử dụng trong nghiên cứu đánh giá đa dạng di truyền kiểu hình quả trên cà chua lưu trữ tại Ngân hàng Gen thực vật quốc gia Iran bởi Marefatzadeh-Khameneh & cs. (2021). Tại Việt Nam, Tran Thien Long & cs. (2020) đã áp dụng phương pháp phân tích thành phần chính và phân tích cụm để xác định các nguồn vật liệu cà chua ưu tú phục vụ chương trình chọn giống cà chua làm cảnh tại Việt Nam.

3.3. Phân tích đa dạng di truyền kiểu hình quả của các vật liệu cà chua bi

Kết quả phân tích cụm phân nhóm đa dạng di truyền của các vật liệu cà chua dựa trên 4 tính trạng quả đo đếm thủ công cho thấy ở khoảng cách di truyền 3,060 (hệ số tương đồng 61,601), 68 nguồn gen cà chua bi được phân thành 6 nhóm chính (Hình 6A, Bảng 4). Sáu nhóm này khác biệt lớn ở khối lượng quả và chiều dài quả (Bảng 5). Các dòng có khối lượng và kích thước quả nhỏ nhất tập trung ở nhóm I gồm 43 dòng, nhóm II gồm 11 dòng, nhóm III gồm 7 dòng, nhóm IV gồm 3 dòng, nhóm V gồm 2 dòng và nhóm VI gồm 2 dòng. Trong đó, khối lượng quả và chiều dài quả trung bình nhóm III < nhóm I < nhóm II < nhóm V < nhóm VI < nhóm IV. Ngoài ra, chiều rộng quả trung bình nhóm III < nhóm I < nhóm II < nhóm V < nhóm IV < nhóm VI. Chỉ số dạng quả nhóm VI < nhóm I < nhóm II-III < nhóm IV < nhóm V. Dựa trên

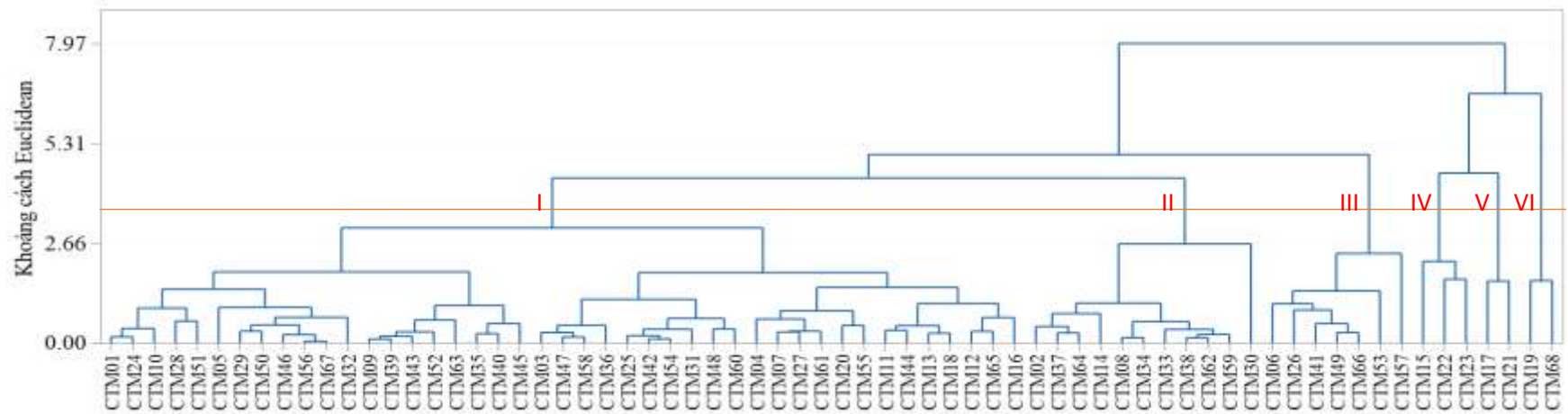
5 tính trạng quả trích xuất qua ảnh, 68 nguồn gen cà chua bi được phân thành 6 nhóm chính ở khoảng cách di truyền 3,566 (hệ số tương đồng 65,945) (Hình 6B, Bảng 4).

Kiểu hình quả là những đặc điểm nông học quan trọng trong các chương trình chọn tạo giống cà chua (Jin & cs., 2019; Zhu & cs., 2022). Đánh giá đa dạng di truyền dựa trên kiểu hình quả cà chua trích xuất qua ảnh đã có nhiều nghiên cứu trên thế giới. Nankar & cs. (2020) đã kết hợp đo đếm thủ công và trích xuất kiểu hình quả qua ảnh để đánh giá đa dạng di truyền của 150 nguồn gen cà chua đại diện cho 21 quốc gia. Marefatzadeh-Khameneh & cs. (2021) đã áp dụng phương pháp trích xuất kiểu hình quả qua ảnh đánh giá đa dạng di truyền 589 nguồn gen cà chua tại Ngân hàng Gen thực vật quốc gia Iran. Ở cả hai nghiên cứu, bằng phương pháp phân tích cụm, nguồn gen cà chua đều được phân thành 10 nhóm khác biệt chính về hình dạng và kích thước quả (Nankar & cs., 2020; Marefatzadeh-Khameneh & cs., 2021). Như vậy, kết quả phân tích đa dạng di truyền dựa trên kiểu hình quả ở nghiên cứu này phù hợp với kết quả của các nghiên cứu trước đó. Phương pháp phân tích này không chỉ áp dụng trên cây cà chua mà còn có thể áp dụng trên các đối tượng cây trồng khác. Tại Việt Nam, Nguyen Trung Duc & cs. (2022) cũng đã áp dụng phương pháp đánh giá kiểu hình quả qua ảnh và sử dụng tính trạng này để đánh giá đa dạng di truyền các nguồn gen dưa lưới.

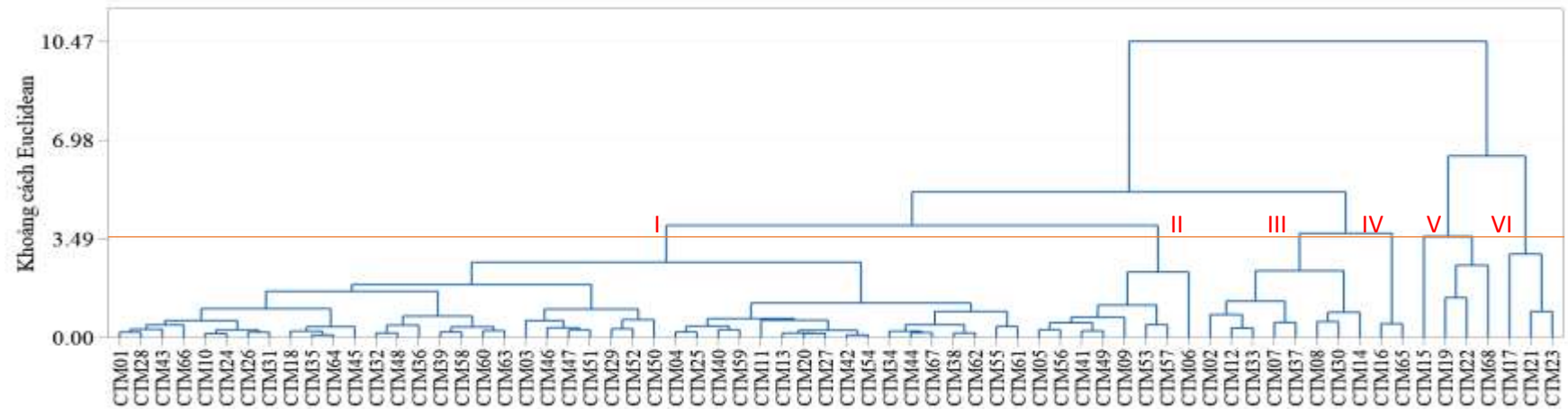


Ghi chú: Dai: Chiều dài quả đo đếm thủ công; Rong: Chiều rộng quả đo đếm thủ công; KLQ: Khối lượng quả đo đếm thủ công; Dientich: Diện tích trích xuất qua ảnh; Dai_img: Chiều dài trích xuất qua ảnh; Rong_img: Chiều rộng trích xuất qua ảnh; LWR_img: Chỉ số dạng quả trích xuất qua ảnh.

Hình 5. Phân tích thành phần chính các tính trạng quả



(A)



(B)

Hình 6. Biểu đồ cụm phân nhóm đa dạng di truyền dựa trên khoảng cách Euclidean từ (B) 04 tính trạng đo đếm thủ công và (B) 05 tính trạng trích xuất qua ảnh

Bảng 4. Kết quả phân nhóm đa dạng di truyền nguồn gen cà chua bi

Phân nhóm dựa trên 4 tính trạng đo đếm thủ công			Phân nhóm dựa trên 5 tính trạng trích xuất qua ảnh		
Số cụm	Hệ số tương đồng	Khoảng cách Euclidean	Số cụm	Hệ số tương đồng	Khoảng cách Euclidean
11	76,537	1,870	11	77,998	2,304
10	76,312	1,888	10	77,534	2,352
9	72,792	2,168	9	75,712	2,543
8	70,178	2,377	8	74,666	2,653
7	66,972	2,632	7	71,906	2,942
6	61,601	3,060	6	65,945	3,566
5	44,976	4,385	5	64,831	3,683
4	43,309	4,518	4	62,212	3,957
3	37,102	5,013	3	50,886	5,143
2	16,805	6,631	2	38,745	6,414
1	0,000	7,970	1	0,000	10,471

Bảng 5. Đặc điểm các nhóm dựa trên kiểu hình đo đếm thủ công

Nhóm	Số dòng	Khối lượng quả (g)		Chiều dài (D) (mm)		Chiều rộng (R) (mm)		Chỉ số dạng quả (D/R)	
		TB	SE	TB	SE	TB	SE	TB	SE
I	43	11,87	0,44	30,66	0,49	27,38	0,46	1,12	0,01
II	11	14,06	0,63	36,96	0,85	28,97	0,76	1,28	0,01
III	7	8,52	0,77	26,56	0,77	20,81	0,87	1,28	0,04
IV	3	39,11	4,44	52,58	1,19	40,22	1,79	1,31	0,05
V	2	27,00	1,00	51,79	4,53	31,64	2,42	1,64	0,01
VI	2	33,81	4,86	39,66	2,31	42,16	0,03	0,95	0,06

Bảng 6. Đặc điểm các nhóm dựa trên kiểu hình trích xuất qua ảnh

Nhóm	Số dòng	Diện tích (mm ²)		Chu vi (mm)		Chiều dài (D) (mm)		Chiều rộng (R) (mm)		Chỉ số dạng quả (D/R)	
		TB	SE	TB	SE	TB	SE	TB	SE	TB	SE
I	43	652,89	13,01	101,29	1,08	30,77	0,32	27,71	0,28	1,11	0,01
II	8	468,21	14,70	85,92	1,79	26,62	0,46	22,55	0,43	1,19	0,02
III	8	765,30	43,92	111,74	3,38	35,94	1,02	28,23	0,75	1,28	0,01
IV	2	1.033,56	30,47	127,62	3,96	38,29	0,50	35,12	0,57	1,09	0,00
V	4	1.645,80	151,76	180,83	14,14	50,57	3,01	42,34	1,54	1,19	0,04
VI	3	1.287,38	116,89	149,82	6,29	49,32	1,43	33,92	2,08	1,46	0,05

4. KẾT LUẬN

Nghiên cứu đánh giá kiểu hình quả nguồn gen cà chua bi đã cho thấy với độ chính xác cao và khả năng tái lập, phương pháp đánh giá kiểu hình quả qua ảnh có thể thay thế phép đo kiểu

hình quả thủ công và ứng dụng trong các chương trình chọn tạo, khảo kiểm nghiệm giống cà chua.

Nguồn gen cà chua thành được phân thành 6 nhóm nhỏ ở hệ số tương đồng 61,601 dựa trên 4 tính trạng đo đếm thủ công và cũng được phân thành 6 nhóm nhỏ ở hệ số tương đồng 65,945 từ 5

tính trạng trích xuất qua ảnh. Có sự tương đồng cao giữa hai kết quả phân nhóm, khác biệt chủ yếu ở kích thước quả. Đây là những nguồn gen ban đầu quan trọng phục vụ chương trình chọn tạo giống cà chua bi ưu thế lai tại Việt Nam.

LỜI CẢM ƠN

Nhóm tác giả trân trọng cảm ơn Đề tài Khoa học và Công nghệ cấp Học viện Nông nghiệp Việt Nam mã số T2021-41-25VB do dự án Việt Bỉ tài trợ đã hỗ trợ một phần kinh phí thực hiện nghiên cứu này.

TÀI LIỆU THAM KHẢO

- Causse M., Friguet C., Coiret C., Lépiciér M., Navez B., Lee M., Holthuysen N., Sinesio F., Moneta E. & Grandillo S. (2010). Consumer preferences for fresh tomato at the European scale: A common segmentation on taste and firmness. *Journal of Food Science*. 75(9): S531-S541.
- Đoàn Xuân Cảnh, Nguyễn Đình Thiệu, Đoàn Thị Thanh Thúy & Nguyễn Thị Thanh Hà (2021). Kết quả nghiên cứu chọn tạo và khảo nghiệm giống cà chua lai VT15. *Tạp chí Nông nghiệp và Phát triển nông thôn*. 9: 34-41.
- He Z., Li M., Cai Z., Zhao R., Hong T., Yang Z. & Zhang Z. (2021). Optimal irrigation and fertilizer amounts based on multi-level fuzzy comprehensive evaluation of yield, growth and fruit quality on cherry tomato. *Agricultural Water Management*. 243: 106360.
- Jin L., Zhao L., Wang Y., Zhou R., Song L., Xu L., Cui X., Li R., Yu W. & Zhao T. (2019). Genetic diversity of 324 cultivated tomato germplasm resources using agronomic traits and InDel markers. *Euphytica*. 215(4): 1-16.
- Kong L., Wen Y., Jiao X., Liu X. & Xu Z. (2021). Interactive regulation of light quality and temperature on cherry tomato growth and photosynthesis. *Environmental and Experimental Botany*. 182: 104326.
- Liu H., Meng F., Miao H., Chen S., Yin T., Hu S., Shao Z., Liu Y., Gao L., Zhu C., Zhang B. & Wang Q. (2018). Effects of postharvest methyl jasmonate treatment on main health-promoting components and volatile organic compounds in cherry tomato fruits. *Food Chemistry*. 263: 194-200.
- Marefatzadeh-Khameneh M., Fabriki-Ourang S., Sorkhilalehloo B., Abbasi-Kohpalekani J. & Ahmadi J. (2021). Genetic diversity in tomato (*Solanum lycopersicum* L.) germplasm using fruit variation implemented by tomato analyzer software based on high throughput phenotyping. *Genetic Resources and Crop Evolution*. 68(6): 2611-2625.
- Nankar A.N., Tringovska I., Grozeva S., Ganeva D. & Kostova D. (2020). Tomato phenotypic diversity determined by combined approaches of conventional and high-throughput tomato analyzer phenotyping. *Plants*. 9(2).
- Nguyễn Hồng Minh, Kiều Thị Thu & Phạm Quang Tuấn (2011). Tạo giống cà chua lai quả nhỏ HT144. *Tạp chí Khoa học và Phát triển*. 9(1): 16-21.
- Nguyen Trung Duc, Pham Quang Tuan, Nguyen Thi Nguyet Anh & Vu Van Liet (2022). Phenotypic variation and correlation of fruit traits in diverse muskmelon materials. *Vietnam Journal of Agriculture & Rural Development*. 2(2): 23-31.
- R Development Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://cran.r-project.org/bin/windows/base> on March 10, 2022.
- Ranc N., Munos S., Santoni S. & Causse M. (2008). A clarified position for *solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (*solanaceae*). *BMC Plant Biology*. 8(1): 130.
- Rodriguez G.R., Moyseenko J.B., Robbins M.D., Morejon N.H., Francis D.M. & Van Der Knaap E. (2010). Tomato Analyzer: a useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. *JoVE (Journal of Visualized Experiments)*. 10.3791/1856(37): e1856.
- Tanabata T., Shibaya T., Hori K., Ebana K. & Yano M. (2012). Smartgrain: High-throughput phenotyping software for measuring seed shape through image analysis *Plant Physiology*. 160(4): 1871-1880.
- Tổng Văn Hải, Phan Hữu Tôn, Phan Thị Hiền & Nguyễn Quốc Trung (2021). Chọn tạo giống cà chua thuần kháng bệnh xoắn vàng lá bằng chỉ thị phân tử ADN. *Tạp chí Khoa học Nông nghiệp Việt Nam*. 19(3): 399-409.
- Tran Thien Long, Nguyen Hong Minh, Nguyen Tuan Anh, Tran Thi Minh Hang, Nguyen Thi Hoa, Nguyen Tien Long & Nguyen Thi Minh (2020). The comprehensive analysis of morphological variation among 24 tomato (*Solanum lycopersicum*) genotypes oriented to ornamental breeding in Vietnam. *Vietnam Journal of Agricultural Sciences*. 3(1): 555-569.
- Zhu Y., Gu Q., Zhao Y., Wan H., Wang R., Zhang X. & Cheng Y. (2022). Quantitative extraction and evaluation of tomato fruit phenotypes based on image recognition. *Frontiers in Plant Science*. 13.