

BỘ GIÁO DỤC VÀ ĐÀO TẠO BỘ NÔNG NGHIỆP VÀ PTNT
VIỆN KHOA HỌC NÔNG NGHIỆP VIỆT NAM

PHẠM THỊ MAI

**NGHIÊN CỨU CHỨC NĂNG CỦA QTL9 LIÊN QUAN ĐẾN
CẤU TRÚC BÔNG, PHỤC VỤ CHỌN TẠO GIỐNG LÚA
NĂNG SUẤT CAO Ở VIỆT NAM**

LUẬN ÁN TIẾN SĨ NÔNG NGHIỆP

Hà Nội, 2022

BỘ GIÁO DỤC VÀ ĐÀO TẠO

BỘ NÔNG NGHIỆP VÀ PTNT

VIỆN KHOA HỌC NÔNG NGHIỆP VIỆT NAM

PHẠM THỊ MAI

**NGHIÊN CỨU CHỨC NĂNG CỦA QTL9 LIÊN QUAN ĐẾN
CẤU TRÚC BÔNG PHỤC VỤ CHỌN TẠO GIỐNG LÚA
NĂNG SUẤT CAO Ở VIỆT NAM**

Chuyên ngành: Công nghệ sinh học

Mã số: 94 20 201

LUẬN ÁN TIẾN SĨ NÔNG NGHIỆP

Người hướng dẫn khoa học:

1. TS. Không Ngân Giang
2. GS.TSKH. Trần Duy Quý

Hà Nội, 2022

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của tôi và nhóm tác giả, các kết quả nghiên cứu được trình bày trong luận án là trung thực, khách quan và chưa từng để bảo vệ ở bất kỳ học vị nào.

Tôi xin cam đoan rằng mọi sự giúp đỡ cho việc thực hiện luận án đã được cảm ơn, các thông tin trích dẫn trong luận án đều có nguồn gốc, xuất xứ rõ ràng.

Hà Nội, ngày tháng năm 2022

Tác giả

Phạm Thị Mai

LỜI CẢM ƠN

Để hoàn thành được luận án, tôi đã nhận được sự quan tâm, giúp đỡ nhiệt tình từ các Thầy, Cô giáo, các tập thể, cá nhân cùng bạn bè, đồng nghiệp và gia đình.

Tôi xin bày tỏ lòng biết ơn sâu sắc tới TS. Không Ngân Giang đã luôn tận tình hướng dẫn, cung cấp tài liệu và tạo điều kiện tốt nhất cho tôi trong quá trình học tập, nghiên cứu thực hiện đề tài và GS.TSKH. Trần Duy Quý đã luôn giúp đỡ, cung cấp những kiến thức quý báu cho tôi trong quá trình hoàn thành luận án này.

Tôi xin chân thành cảm ơn sự quan tâm tạo điều kiện của Lãnh đạo Viện Di truyền Nông nghiệp, Phòng Thí nghiệm Trọng điểm Công nghệ tế bào thực vật đã giúp đỡ, tạo điều kiện thuận lợi cho tôi trong quá trình học tập, triển khai đề tài.

Tôi xin chân thành cảm ơn tập thể các Thầy, Cô và các cán bộ công tác tại Ban Đào tạo Sau đại học – Viện Khoa học Nông nghiệp Việt Nam đã giảng dạy và tạo điều kiện thuận lợi cho tôi trong suốt quá trình học tập tại Viện.

Tôi xin chân thành cảm ơn ThS. Lê Thị Như, ThS. Trần Vũ Hằng, ThS. Vũ Thị Nhiên (Viện Di truyền Nông nghiệp), TS. Stefan Juannic (Viện Nghiên cứu và Phát triển IRD - Pháp) đã giúp đỡ, động viên và đồng hành cùng tôi trong quá trình thực hiện luận án.

Để hoàn thành được luận án, không thể thiếu sự động viên, khuyến khích, tạo điều kiện của gia đình, đó là nguồn động lực lớn giúp tôi hoàn thành luận án này.

Tôi xin chân thành cảm ơn!

Hà Nội, ngày tháng năm 2022

Tác giả

Phạm Thị Mai

MỤC LỤC

Lời cam đoan.....	i
Lời cảm ơn.....	ii
Mục lục.....	iii
Danh mục ký hiệu và chữ viết tắt.....	vi
Danh mục bảng.....	viii
Danh mục hình	ix
Mở đầu	1
1. Tính cấp thiết của luận án	1
2. Mục tiêu của luận án	2
3. Ý nghĩa khoa học và thực tiễn của đề tài	3
4. Đối tượng và phạm vi nghiên cứu của đề tài	3
5. Tính mới và những đóng góp của luận án.....	5
Chương I. Tổng quan tài liệu và cơ sở khoa học của đề tài	6
1.1. Tầm quan trọng của cây lúa	6
1.2. Tính trạng năng suất và các yếu tố cấu thành năng suất lúa.....	7
1.3. QTL và lập bản đồ QTL.....	10
1.4. Các QTL liên kết với tính trạng năng suất và các yếu tố cấu thành năng suất ở lúa.....	12
1.5. Các quần thể lập bản đồ QTL liên quan đến tính trạng năng suất lúa.....	14
1.6. Cấu trúc bông lúa và các QTL/gen liên quan đến cấu trúc bông lúa.....	19
1.7. Chỉ thị phân tử CAPS và ứng dụng trong nghiên cứu lập bản đồ QTL liên quan đến tính trạng năng suất lúa.....	24
1.8. Đa hình nucleotit đơn (Single Nucleotide Polymorphism - SNP).....	26
1.9. Tìm kiếm SNPs trong vùng genom mục tiêu bằng công nghệ chụp gen kết hợp với giải trình tự thế hệ mới.....	28

1.10. Nghiên cứu liên kết trên toàn hệ gen (GWAS), tiềm năng ứng dụng trong nghiên cứu về các tính trạng nông học phức tạp ở lúa	30
1.11. Haplotype và phương pháp phân tích haplotype	34
Chương II. Vật liệu, nội dung và phương pháp nghiên cứu	36
2.1. Vật liệu và các thiết bị trong nghiên cứu	36
2.2. Địa điểm và thời gian nghiên cứu	39
2.3. Nội dung nghiên cứu	39
2.4. Phương pháp nghiên cứu.....	40
2.4.1. Phân tích haplotype vùng QTL9 của tập đoàn lúa sử dụng trong nghiên cứu GWAS và xác định các giống lúa bố mẹ để tạo quần thể lai F ₁	40
2.4.2. Phương pháp tạo các quần thể lai	41
2.4.3. Chọn lọc cây F ₁ bằng chỉ thị phân tử SSR	42
2.4.4. Xác định các SNPs trong vùng QTL9 ở các giống bố mẹ bằng phương pháp chụp gen (Gene capture) kết hợp với giải trình tự thế hệ mới	43
2.4.5. Phát triển chỉ thị phân tử CAPS	48
2.4.6. Chọn lọc các cây F ₂ mang QTL9 đồng hợp thuộc hai haplotype bố hoặc mẹ bằng chỉ thị phân tử CAPS	49
2.4.7. Phương pháp bố trí thí nghiệm và các chỉ tiêu theo dõi ngoài đồng ruộng.....	50
Chương III. Kết quả nghiên cứu và thảo luận	54
3.1. Chọn lọc các cặp lai bố mẹ và lai tạo quần thể F ₁	54
3.1.1. Phân tích haplotype vùng QTL9 của tập đoàn lúa sử dụng trong nghiên cứu GWAS, chọn lọc các giống lúa bố mẹ làm vật liệu lai tạo	54
3.1.2. Tạo quần thể lai F ₁ và chọn lọc cây F ₁ bằng chỉ thị phân tử SSR..	60

3.2. Phát triển chỉ thị phân tử CAPS để chọn lọc các cây F ₂ mang QTL9 đồng hợp thuộc hai haplotype bố hoặc mẹ.....	69
3.2.1. Xác định các SNPs trong vùng QTL9 của hai giống bố mẹ để phát triển chỉ thị phân tử CAPS	69
3.2.2. Phát triển chỉ thị phân tử CAPS	83
3.3. Chọn lọc các cây F ₂ mang QTL9 đồng hợp thuộc hai haplotype bố hoặc mẹ bằng chỉ thị phân tử CAPS	86
3.3.1. Xác định chỉ thị phân tử CAPS cho đa hình giữa hai giống lúa bố mẹ.....	86
3.3.2. Chọn lọc các cây F ₂ đồng hợp tử mang QTL9 của bố hoặc mẹ bằng chỉ thị phân tử CAPS	91
3.4. Phân tích kiểu hình cấu trúc bông của hai quần thể F ₃ thuộc hai haplotype bố hoặc mẹ.....	97
Kết luận và kiến nghị	112
1. Kết luận	112
2. Kiến nghị.....	113
Danh mục công trình đã công bố liên quan đến luận án.....	114
Tài liệu tham khảo.....	115
Phụ lục	

DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Thứ tự	Chữ viết tắt	Nghĩa của chữ viết tắt
1	AFLP (<i>Amplified Fragment Length Polymorphism</i>)	Đa hình chiều dài đoạn nhân bản khuếch đại
2	BILs (<i>Backcross Recombinant Inbred Lines</i>)	Quần thể lai trở lại cận giao tái tổ hợp
3	Bp (<i>Base pair</i>)	Cặp bazơ nitơ
4	CAPS (<i>Cleaved amplified polymorphic sequences</i>)	Đa hình khuếch đại đoạn phân cắt
5	DHs (<i>Doubled haploids</i>)	Đơn bội kép
6	DNA	Deoxyribonucleic acid
7	GBS (<i>Genotyping by sequencing</i>)	Xác định kiểu gen bằng giải trình tự
8	GWAS (<i>Genome wide association study</i>)	Nghiên cứu liên kết trên toàn hệ gen
9	H1	Haplotype 1
10	H2	Haplotype 2
11	Kb	Kilo bazơ nitơ
12	LD (<i>Linkage Disequilibrium</i>)	Mất cân bằng di truyền liên kết
13	MAS (<i>Marker assisted selection</i>)	Chọn lọc nhờ chỉ thị phân tử
14	NGS (<i>Next generation sequencing</i>)	Giải trình tự thế hệ mới
15	NST	Nhiễm sắc thể
16	PBintL (<i>Primary branch internode average length</i>)	Khoảng cách giữa các gié cấp một

Thứ tự	Chữ viết tắt	Nghĩa của chữ viết tắt
17	PBN (<i>Primary branch number</i>)	Số gié cấp một/bông
18	PBL (<i>Primary branch length</i>)	Chiều dài gié cấp một
19	PCR (<i>Polymerase chain reaction</i>)	Phản ứng chuỗi polymerase
20	QTL (<i>Quantitative trait loci</i>)	Lô-cut tính trạng số lượng
21	RAPD (<i>Random Amplified Polymorphic DNA</i>)	Đa hình khuếch đại các đoạn ngẫu nhiên DNA
22	RFLP (<i>Restriction Fragment Length Polymorphism</i>)	Đa hình chiều dài đoạn cắt giới hạn
23	RILs (<i>Recombinant Inbred Lines</i>)	Dòng lai cận giao tái tổ hợp
24	RL (<i>Rachis length</i>)	Chiều dài bông
25	SBN (<i>Secondary branch number</i>)	Số gié cấp hai/bông
26	SBintL (<i>Secondary branch internode average length</i>)	Khoảng cách giữa các gié cấp hai
27	SBL (<i>Secondary branch length</i>)	Chiều dài gié cấp hai
28	SNP (<i>Single nucleotide polymorphism</i>)	Đa hình nucleotide đơn
29	SpN (<i>Spikelet number</i>)	Số hạt/bông
30	SSR (<i>Simple Sequence Repeat</i>)	Lặp lại trình tự đơn giản
31	TBN (<i>Tertiary branch number</i>)	Số gié cấp ba/bông

DANH MỤC BẢNG

Bảng	Tên bảng	Trang
2.1	Một số chỉ số về cấu trúc bông và năng suất của các giống lúa sử dụng làm bố mẹ để tạo các quần thể lai F ₁	37
2.2	Danh sách 12 chỉ thị phân tử SSR sử dụng để chọn lọc cây F ₁ ...	38
3.1	Danh sách các giống lúa thuộc 9 haplotype và trình tự haplotype	55
3.2	Các cặp lai được tiến hành từ 4 giống bố mẹ.....	60
3.3	Kết quả giống hàng	70
3.4	Kết quả tìm và sàng lọc biến thể.....	72
3.5	Phân loại biến thể trong vùng exon.....	73
3.6	Chú giải biến thể đồng nghĩa giữa hai haplotype	74
3.7	Chú giải biến thể sai nghĩa giữa hai haplotype	76
3.8	Chú giải biến thể đột biến khung đọc giữa hai haplotype	79
3.9	Chú giải biến thể mất bộ ba mã hóa.....	80
3.10	Chú giải biến thể thêm bộ ba mã mở đầu và thêm bộ ba mã kết thúc	81
3.11	Danh sách SNP nằm trong vị trí cắt của các enzyme giới hạn	82
3.12	Danh sách 12 chỉ thị phân tử CAPS và trình tự môi.....	84
3.13	Trình tự vị trí cắt và kích thước các sản phẩm cắt của 5 enzyme giới hạn sử dụng để cắt chỉ thị phân tử CAPS.....	85
3.14	Danh sách cây F ₂ đồng hợp tử được xác định bằng 3 chỉ thị phân tử CAPS 1, CAPS 5, CAPS 11.....	95

DANH MỤC HÌNH

Hình	Tên hình	Trang
1	Phân tích GWAS tính trạng số hạt/bông (SpN) và số gié cấp hai/bông (SBN) của tập đoàn lúa bản địa Việt Nam.....	4
1.1	Bản đồ vị trí các QTL liên kết với tính trạng năng suất trên 12 nhiễm sắc thể ở lúa.....	13
1.2	Cấu trúc bông lúa (A) và sự thay đổi qua quá trình thuần hóa (B) ...	20
1.3	Các giai đoạn hình thành bông lúa non quan sát dưới kính hiển vi ..	21
2.1	Sơ đồ lai và chọn lọc MAS để đánh giá vai trò của QTL9 liên quan đến cấu trúc bông lúa.....	40
2.2	Quá trình lai mẫu dò và DNA vùng mục tiêu (theo MyBaits V2 của MYcroarray)	45
2.3	Các bước gọi biến thể sử dụng TOGGLE	46
2.4	Các bước sàng lọc biến thể và các tiêu chí sàng lọc tương ứng với từng bước	47
2.5	Mô hình sử dụng chỉ thị CAPS xác định cá thể đồng hợp tử	50
2.6	Mô hình bố trí các ô thí nghiệm ngoài đồng ruộng.....	51
2.7	Cấu trúc bông lúa (A) và đưa vào phần mềm phân tích P-TRAP (B)	52
3.1	Phân tích haplotype vùng QTL9	54
3.2	Phân tích cấu trúc bông của 2 haplotype đại diện (H1, H2)	57
3.3	Hình thái cấu trúc bông của 4 giống lúa sử dụng làm bố, mẹ để tạo quần thể lai F ₁	58
3.4	Quá trình xử lý chu kỳ quang và lai giữa các giống lúa bố mẹ để tạo quần thể F ₁	61
3.5	Kết quả khảo sát sự đa hình của các chỉ thị phân tử SSR ở các	62

Hình	Tên hình	Trang
	giống bố mẹ (G6, G19, G189, G205)	
3.6	Kết quả chọn lọc cây F ₁ của cặp lai G6 x G189 với ba chỉ thị phân tử RM320 (A), RM180 (B) và RM491 (C).....	65
3.7	Kết quả chọn lọc cây F ₁ của cặp lai G6 x G205 với hai chỉ thị phân tử RM320 (A), RM180 (B)..	66
3.8	Hình. Kết quả chọn lọc cây F ₁ của cặp lai G19 x G189 với ba chỉ thị phân tử RM289 (A), RM592 (B) và RM204 (C).....	67
3.9	Kết quả chọn lọc cây F ₁ của cặp lai G19 x G205 với ba chỉ thị phân tử RM289 (A), RM592 (B) và RM204 (C).....	68
3.10	Kết quả phản ứng PCR thử độ đặc hiệu của các môi CAPS ở hai giống G6 (H1) và G189 (H2)	87
3.11	Kết quả điện di sản phẩm cắt bằng enzyme giới hạn.....	89
3.12	Kết quả phân tích một số cây F ₂ với chỉ thị CAPS 1	92
3.13	Kết quả phân tích một số cây F ₂ với chỉ thị CAPS 5	93
3.14	Kết quả phân tích một số cây F ₂ với chỉ thị CAPS 11	94
3.15	Phân tích định lượng các tính trạng cấu trúc bông lúa.....	99
3.16	Biểu đồ Q-Q plot về sự phân bố của các tính trạng cấu trúc bông lúa.....	102
3.17	Biểu đồ hộp so sánh 8 tính trạng cấu trúc bông của quần thể F ₃ đồng hợp tử mang QTL9 của mẹ (F ₃ _H1) và bố (F ₃ _H2).....	103
3.18	Biểu đồ hộp so sánh bốn tính trạng liên quan đến năng suất của quần thể F ₃ đồng hợp tử mang QTL9 của mẹ (F ₃ _H1) và bố (F ₃ _H2)	104
3.19	Biểu đồ phân tích tương quan thành phần chính PCA.....	105
3.20	Biểu đồ mối tương quan Corrplot giữa các tính trạng cấu trúc	

Hình	Tên hình	Trang
	bông ở quần thể F_3	106
3.21	Mô hình hồi quy tuyến tính đơn biến giữa hai tính trạng số gié cấp hai/bông (SBN) và số hạt/bông (SpN) (A) và tương quan tuyến tính giữa các tính trạng SpN và SBN trong hai quần thể con F_3 thuộc hai haplotype (B)	107
3.22	So sánh cấu trúc bông của hai haplotype.....	109

MỞ ĐẦU

1. Tính cấp thiết của luận án

Lúa (*Oryza sativa*. L) là một trong những cây trồng quan trọng hàng đầu cung cấp lương thực cho hơn 50% dân số thế giới [95]. Để đáp ứng nhu cầu tiêu thụ do bùng nổ dân số, đô thị hoá và những ảnh hưởng của biến đổi khí hậu, dự đoán đến năm 2040, sản lượng gạo cần tăng thêm ít nhất 20% [25]. Là quốc gia xuất khẩu gạo lớn thứ hai thế giới, sản lượng lúa gạo của Việt Nam đóng vai trò quan trọng đối với an ninh lương thực ở khu vực Châu Á. Trong những năm gần đây, năng suất lúa được cải thiện nhưng diện tích sản xuất lúa ngày càng giảm do quá trình đô thị hóa, chuyển đổi cơ cấu cây trồng khác thay thế cho cây lúa diễn ra mạnh mẽ nên sản lượng lúa gạo vẫn chưa đáp ứng được nhu cầu tăng dân số toàn cầu [139]. Do đó, việc nâng cao năng suất lúa gạo trở thành mục tiêu quan trọng đối với các quốc gia trồng lúa trong đó có Việt Nam.

Năng suất là tính trạng phức tạp được quy định bởi nhiều gen. Để đáp ứng mục tiêu tăng năng suất lúa thì những hiểu biết về các vùng genom quy tụ nhiều gen hay còn gọi là QTL (Quantitative Trait Loci - Lô-cut tính trạng số lượng) liên quan đến năng suất là vô cùng cần thiết. Năng suất lúa được xác định bởi ba tính trạng thành phần chính: Số lượng bông, số lượng hạt trên bông và khối lượng hạt [116]. Cấu trúc bông lúa là một tính trạng nông học quan trọng trực tiếp ảnh hưởng đến năng suất [30] và việc đạt được cấu trúc, kích thước, hình dạng bông lúa tối ưu là một trong những mục tiêu chọn tạo giống lúa năng suất cao [94]. Hiểu các cơ chế di truyền cơ bản kiểm soát sự phát triển bông lúa có ý nghĩa rất lớn trong việc cải thiện năng suất lúa gạo [65].

Trong nghiên cứu trước đây của Khổng Ngân Giang, Tạ Kim Nhung và các cộng sự, Viện Di truyền Nông nghiệp đã ứng dụng phương pháp nghiên cứu liên kết trên toàn hệ gen (Genome Wide Association Study - GWAS) để

phân tích tính trạng năng suất của tập đoàn lúa bản địa Việt Nam, đã xác định được 29 QTLs tiềm năng liên quan đến cấu trúc bông lúa, trong đó có một QTL hoàn toàn mới được đặt tên là QTL9 liên kết với cả hai tính trạng số gié cấp hai/bông và số hạt/bông là các tính trạng quan trọng ảnh hưởng đến năng suất lúa. QTL9 nằm trên nhiễm sắc thể số 2, có chiều dài 780 kb, 137 gen được tìm thấy trong vùng QTL này nhưng chưa gen nào được nghiên cứu chức năng [102]. Tuy nhiên, kết quả phân tích GWAS dựa trên các mô hình phân tích thống kê, để có thể khai thác sử dụng vào các chương trình chọn tạo giống, các QTLs thu được từ phân tích GWAS cần phải được nghiên cứu chức năng thông qua các quần thể lai và phát triển các chỉ thị phân tử. Do đó chúng tôi tiến hành thực hiện đề tài “*Nghiên cứu chức năng của QTL9 liên quan đến cấu trúc bông, phục vụ chọn tạo giống lúa năng suất cao ở Việt Nam*”.

2. Mục tiêu của luận án

2.1. Mục tiêu chung

Nghiên cứu xác định vai trò của QTL9 liên quan đến cấu trúc bông lúa thu được từ phân tích GWAS trên tập đoàn lúa bản địa Việt Nam nhằm ứng dụng trong các chương trình chọn tạo giống lúa năng suất cao ở Việt Nam.

2.2. Mục tiêu cụ thể

- Chọn lọc được các cặp lai bố mẹ mang QTL9 thuộc hai haplotype tương phản về cấu trúc bông (bông to và bông nhỏ) và lai tạo được các quần thể lai tái tổ hợp (F_1 , F_2 , F_3) làm vật liệu nghiên cứu.

- Phát triển được bộ chỉ thị phân tử CAPS bao phủ vùng QTL9 cho phép phân biệt kiểu gen QTL9 ở hai haplotype bố mẹ, nhằm chọn lọc các cây F_2 mang QTL9 đồng hợp tử thuộc hai haplotype bố hoặc mẹ.

- Xác định được vai trò của QTL9 liên quan đến cấu trúc bông phục vụ công tác chọn tạo giống lúa năng suất cao ở Việt Nam.

3. Ý nghĩa khoa học và thực tiễn của đề tài

3.1. Ý nghĩa khoa học

- Cung cấp các dẫn liệu khoa học có giá trị về vai trò của QTL9 liên quan đến cấu trúc bông lúa, cụ thể là hai tính trạng số gié cấp hai/bông và số hạt/bông, cũng như phương pháp đánh giá kiểu gen, kiểu hình của các quần thể lai trong việc đánh giá vai trò của QTL nghiên cứu.

- Các kết quả nghiên cứu của đề tài cũng là tài liệu tham khảo trong việc nghiên cứu, giảng dạy có giá trị trong lĩnh vực công nghệ sinh học và di truyền chọn giống cây trồng nói chung và cây lúa nói riêng.

3.2. Ý nghĩa thực tiễn

- Kết quả nghiên cứu của luận án đã xác định được vai trò của QTL9 có tham gia vào việc quy định kiểu hình cấu trúc bông, đặc biệt là hai tính trạng số gié cấp hai/bông và số hạt/bông, từ đó ảnh hưởng đến năng suất lúa, là công cụ phục vụ công tác chọn tạo giống lúa năng suất cao.

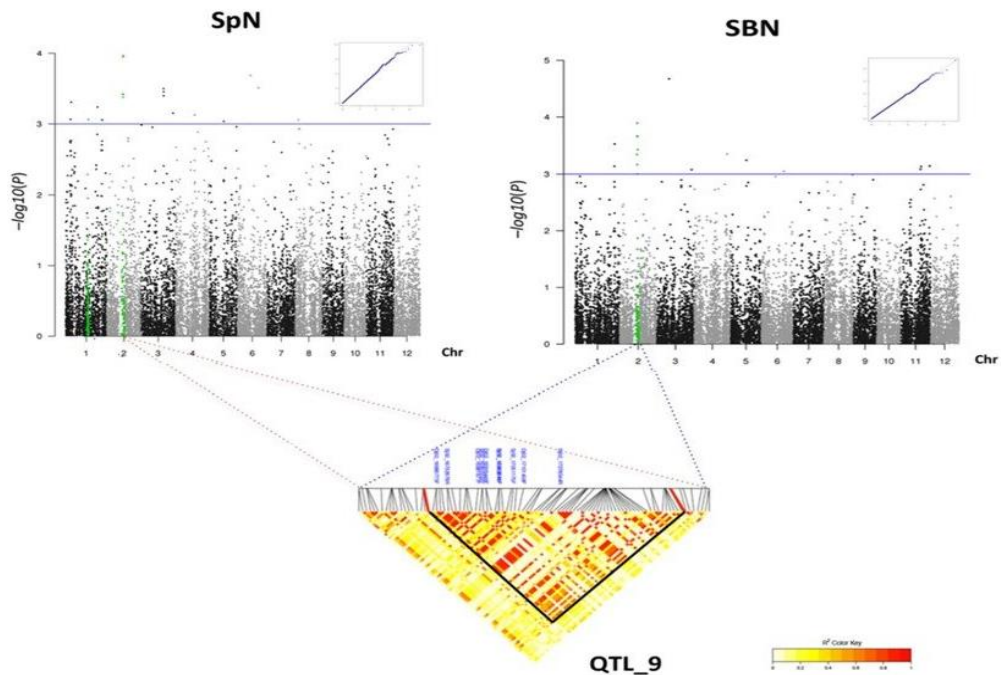
- Kết quả luận án cung cấp bộ chỉ thị phân tử liên kết với QTL9, hỗ trợ các nhà khoa học trong chọn tạo giống lúa năng suất cao nhờ chỉ thị phân tử.

- Cung cấp các giống lúa bố mẹ sử dụng làm vật liệu trong lai tạo để cải tạo năng suất lúa.

4. Đối tượng và phạm vi nghiên cứu của đề tài

4.1. Đối tượng nghiên cứu

Là QTL9 ở cây lúa phát hiện từ nghiên cứu liên kết trên toàn hệ gen (GWAS) trên tập đoàn lúa bản địa Việt Nam do nhóm tác giả Khổng Ngân Giang, Tạ Kim Nhung và cộng sự nghiên cứu [102]. QTL9 dài 780 kb, chứa 9 SNP (gagagcgaa/atataaatt), nằm trên nhiễm sắc thể số 2, liên kết với hai tính trạng số gié cấp hai/bông (SBN) và số hạt/bông (SpN) (Hình 1).



Hình 1. Phân tích GWAS tính trạng số hạt/bông (SpN) và số gié cấp hai/bông (SBN) của tập đoàn lúa bản địa Việt Nam

Từ trên xuống dưới QQ plot, Manhattan plot của 12 nhiễm sắc thể cho 2 tính trạng số hạt/bông (SpN), số gié cấp hai/bông (SBN) và mật cân bằng di truyền liên kết (Linkage Disequilibrium (LD) heatmap) xung quanh vùng QTL9.

Nguồn: Tạ Kim Nhung và cộng sự (2018) [102].

4.2. Địa điểm và thời gian nghiên cứu

- Địa điểm thực hiện:

+ Thí nghiệm phân tích, đánh giá kiểu gen QTL9 các quần thể F_1 , F_2 , F_3 thực hiện tại Viện Di truyền Nông nghiệp (phường Cổ Nhuế 1, quận Bắc Từ Liêm, thành phố Hà Nội);

+ Thí nghiệm chụp gen kết hợp giải trình tự thế hệ mới Illumina và phân tích SNP, phát triển chỉ thị phân tử CAPS thực hiện tại Viện Nghiên cứu và phát triển IRD - Pháp;

+ Thí nghiệm đánh giá kiểu hình thực hiện ngoài đồng ruộng tại xã Đông La, huyện Hoài Đức, thành phố Hà Nội.

- Thời gian thực hiện luận án: Từ năm 2017 đến năm 2020.

5. Tính mới và những đóng góp của luận án

- Đây là một trong những công trình đầu tiên tại Việt Nam nghiên cứu xác định vai trò của một QTL mới (QTL9) phát hiện từ phân tích GWAS trên tập đoàn lúa bản địa Việt Nam liên kết với hai tính trạng số gié cấp hai/bông và số hạt/bông.

- Phân tích haplotype của QTL9 đã xác định được 4 giống lúa bố mẹ có cấu trúc bông khác biệt (bông to và bông nhỏ), thuộc hai haplotype, có khoảng cách di truyền xa nhau. Các giống lúa này có thể sử dụng làm vật liệu cho các nghiên cứu chọn tạo giống lúa năng suất cao ở Việt Nam.

- Sử dụng công nghệ chụp gen (Gene Capture) và giải trình tự thế hệ mới Illumina đã phát hiện 12 SNPs nằm trong vùng QTL9 từ đó phát triển được 12 chỉ thị phân tử CAPS (Cleaved Amplified Polymorphic Sequences), trong đó, 3 chỉ thị phân tử CAPS cho đa hình giữa hai giống lúa bố mẹ trong nghiên cứu. Các chỉ thị phân tử CAPS này cho phép chọn lọc các dòng đồng hợp tử mang kiểu gen QTL9 của bố hoặc của mẹ ngay ở thế hệ thứ 2 thay vì 8 - 10 thế hệ khi sử dụng phương pháp chọn lọc truyền thống.

- Luận án đã khẳng định được vai trò của QTL9 có tham gia vào sự hình thành cấu trúc bông lúa từ đó QTL9 có thể sử dụng làm vật liệu trong lai tạo nhằm cải thiện năng suất cho các giống lúa thương mại có năng suất thấp, chất lượng cao tại Việt Nam.

CHƯƠNG I

TỔNG QUAN TÀI LIỆU VÀ CƠ SỞ KHOA HỌC CỦA ĐỀ TÀI

1.1. Tầm quan trọng của cây lúa

Lúa là một trong những cây ngũ cốc quan trọng nhất về lượng calo và dinh dưỡng cho con người [25], [122]. Để đáp ứng nhu cầu lương thực của người dân toàn cầu, sản lượng nông nghiệp tăng hơn 50% vào năm 2050 [81], đặc biệt là các nước đang phát triển ở Châu Á, Châu Phi. Lúa gạo là thực phẩm chính của hơn một nửa dân số thế giới và cung cấp hơn 20% tổng năng lượng hấp thụ hàng ngày của con người [133]. Đối với một số quốc gia đang phát triển, nguồn lợi từ việc xuất khẩu lúa gạo đóng vai trò quan trọng trong cơ cấu kinh tế quốc gia trong đó có Việt Nam. Là nước xuất khẩu gạo đứng thứ hai thế giới, đồng thời cũng là một trong những nước có nghề truyền thống trồng lúa nước cổ xưa nhất thế giới, tại Việt Nam, cây lúa không chỉ đảm bảo an ninh lương thực quốc gia, mà còn đóng vai trò quan trọng trong cơ cấu cây trồng và phân công lao động xã hội [132].

Hiện nay, Tổ chức Lương thực và Nông nghiệp Liên hợp quốc (FAO) cảnh báo, những diễn biến khó lường của biến đổi khí hậu và thời tiết, dân số tăng nhanh, tác động tiêu cực của dịch bệnh và xung đột liên miên tại nhiều quốc gia đang khiến tình hình an ninh lương thực toàn cầu xấu đi. Mặc dù sản lượng lúa gạo tăng đều hàng năm nhưng tốc độ tăng sản lượng vẫn không đủ đáp ứng nhu cầu của thế giới. Trong khi đó diện tích đất canh tác giảm mạnh do quá trình đô thị hóa, chất lượng đất ngày càng xấu đi, sử dụng phân bón không hợp lý phá hỏng hệ sinh thái và môi trường, một bộ phận lực lượng lao động chuyển sang các ngành công nghiệp, chi phí sản xuất lúa gạo ngày càng tăng cao. Trong giai đoạn 2011 - 2013, thế giới có khoảng 842 triệu người thiếu lương thực, chiếm 12% dân số thế giới. Châu Phi có nguy cơ bị mất an ninh lương thực nhất thế giới, với 29/38 quốc gia bị liệt vào danh sách này [132].

Để giải quyết các vấn đề an ninh lương thực, có nhiều giải pháp được đưa ra như mở rộng đất canh tác, sử dụng chất kích thích tăng năng suất ... Tuy nhiên các nhà khoa học đặt quan tâm nhiều đến việc tạo ra các giống lúa năng suất cao, chất lượng tốt, có khả năng chống chịu các điều kiện bất lợi. Sự phát triển các kỹ thuật di truyền cho phép các nhà khoa học tập trung cải thiện và quy tụ các tính trạng tốt về cùng một giống lúa. Vì vậy, việc nghiên cứu các QTL liên quan đến tính trạng năng suất sẽ góp phần cải tạo năng suất lúa và ứng dụng trong chương trình chọn tạo giống lúa năng suất cao.

1.2. Tính trạng năng suất và các yếu tố cấu thành năng suất lúa

Năng suất là một trong những tính trạng nông học phức tạp, được xác định bởi các yếu tố trực tiếp như: số bông trên cây, số hạt trên bông và khối lượng hạt [52], [116] và một số các yếu tố tác động gián tiếp như: chiều cao cây, thời gian ra hoa, số nhánh hữu hiệu, số lá, khả năng quang hợp... [37], [85], [86]. Trong khi khối lượng hạt được xác định bởi hai thành phần là kích thước hạt (chiều dài, chiều rộng và độ dày hạt) và tỉ lệ hạt chắc, số lượng bông phụ thuộc vào khả năng đẻ nhánh thì số lượng hạt/bông phụ thuộc vào cấu trúc bông bao gồm rất nhiều các yếu tố cấu thành: số lượng gié cấp một, gié cấp hai, gié cấp ba và cuối cùng là số lượng hoa. Ở lúa, mỗi một hoa sẽ phát triển thành một hạt, do đó số lượng hoa quy định số lượng hạt trên bông [52], [113].

Năng suất là một tính trạng mang tính quần thể vì trong thực tế sản xuất tính trạng này luôn được so sánh trên một diện tích sản xuất với hàng nghìn cây chứ không chỉ trên một cá thể đơn lẻ, song nghiên cứu trên quy mô nhỏ sẽ là bước khởi đầu để khảo sát đại trà trên diện tích lớn hơn. Để tăng năng suất, các nghiên cứu thường tập trung vào tăng số lượng hạt lúa trên một bông và số bông trên cây [1].

Theo Nguyễn Văn Hoan (2006), sự tương quan giữa năng suất và số bông/cây ở mỗi giống lúa là khác nhau, ở những giống thuộc nhóm bán lùn có sự tương quan chặt ($r = 0,85$), nhóm lùn có tương quan vừa ($r = 0,62$) và nhóm cao cây có tương quan vừa ($r = 0,54$). Sự tương quan giữa năng suất và số hạt/bông thì ngược lại, nhóm cao cây có tương quan rất chặt ($r = 0,96$), nhóm bán lùn và lùn có tương quan vừa ($r = 0,62 - 0,66$). Mối quan hệ giữa các yếu tố cấu thành năng suất, năng suất thực thu thực chất là mối quan hệ giữa cá thể và quần thể. Khi mật độ và số bông/m² tăng trong phạm vi nào đó khối lượng bông giảm ít thì năng suất cuối cùng tăng. Nhưng nếu số bông/m² tăng cao quá sẽ làm khối lượng bông giảm nhiều, lúc đó năng suất sẽ giảm. Vì vậy cần phải điều tiết mối quan hệ này sao cho hợp lý để năng suất cuối cùng đạt cao nhất [5].

Số hạt/bông nhiều hay ít tùy thuộc vào số gié bao gồm các gié cấp một, gié cấp hai, gié cấp ba, số hoa phân hóa cũng như thoái hóa, tỷ lệ tạo hạt của bông con. Toàn bộ quá trình này nằm trong thời kỳ sinh trưởng sinh thực (từ làm đòng đến trổ). Và số lượng gié, hoa phân hóa được quyết định ngay từ thời kỳ đầu của quá trình làm đòng (bước 1 - 3 trong vòng từ 7 - 10 ngày). Thời kỳ này bị ảnh hưởng bởi sinh trưởng của cây lúa và điều kiện ngoại cảnh, các yếu tố này cũng ảnh hưởng trực tiếp đến sự thoái hóa hoa. Thời kỳ thoái hóa hoa thường bắt đầu vào bước 4 (hình thành nhị và nhụy) và kết thúc vào bước 6, tức là khoảng 10 - 12 ngày trước trổ. Nguyên nhân chủ yếu do thiếu dinh dưỡng ở thời kỳ làm đòng hoặc do ngoại cảnh bất thuận như trời rét, âm u, thiếu ánh sáng, bị ngập, hạn, sâu bệnh... ngoài ra cũng có nguyên nhân do đặc điểm của một số giống [2], [3]. Số hoa phân hóa càng nhiều, số hoa thoái hóa càng ít thì số hạt/bông càng lớn. Tỷ lệ hoa phân hóa có liên quan chặt chẽ với chế độ chăm sóc. Số gié cấp một, đặc biệt là số gié cấp hai

nhieu thì số hoa/bông nhiều, đây sẽ là điều kiện cần thiết để đảm bảo cho tổng hợp số hạt/bông lớn.

Tỉ lệ hạt chắc/bông: tăng tỉ lệ hạt chắc/bông hay nói cách khác là giảm tỉ lệ hạt lép/bông cũng là yếu tố quan trọng quyết định năng suất lúa. Tỉ lệ hạt chắc/bông được quyết định ở thời kỳ trước và sau trổ, nếu gặp điều kiện bất thuận trong thời kỳ này thì tỉ lệ lép sẽ cao. Tỉ lệ lép/bông không chỉ bị ảnh hưởng của các yếu tố nói trên mà còn bị ảnh hưởng bởi đặc điểm của giống. Thường tỉ lệ lép dao động tương đối lớn, trung bình từ 5 - 10%, ít là 2 - 5%, cũng có khi trên 30% hoặc thậm chí còn cao hơn nữa [2], [3].

Yếu tố cuối cùng là khối lượng nghìn hạt: yếu tố này biến động không nhiều do điều kiện dinh dưỡng và ngoại cảnh, chủ yếu phụ thuộc vào yếu tố giống. Khối lượng nghìn hạt được cấu thành bởi hai yếu tố: khối lượng vỏ trấu (thường chiếm khoảng 20%) và khối lượng hạt gạo (thường chiếm khoảng 80%). Vì vậy muốn khối lượng hạt gạo cao, phải tác động vào cả hai yếu tố này. Khối lượng nghìn hạt được xác định phần lớn do kích thước hạt, được quy định bởi ba yếu tố: Chiều dài, chiều rộng và độ dày [90], [112]. Để tăng khối lượng hạt, trước lúc trổ bông, cần bón nuôi đòng để làm tăng kích thước vỏ trấu. Sau khi trổ bông cần tạo điều kiện cho cây sinh trưởng tốt để quang hợp được tiến hành mạnh mẽ, tích lũy được nhiều tinh bột, khối lượng hạt sẽ cao.

Chiều dài bông là một đặc điểm di truyền của giống, nó được tính từ đót cổ bông đến đầu mút bông. Chiều dài bông là một tính trạng liên quan trực tiếp đến số hạt/bông, nó quyết định một phần năng suất của giống. Chiều dài bông do cả gen trội và gen lặn quy định. Chiều dài cổ bông do các gen trội điều khiển và có độ biến động rất lớn, có liên quan đến chiều dài lóng đót cuối cùng và biểu hiện ở tính trổ của bông. Trong nghiên cứu lúa lai, các nhà khoa học đã phát hiện gen lặn có khả năng kéo dài lóng đót cuối cùng mạnh

nhất làm cỏ bông dài ra nhưng không kéo dài các lóng ở bên dưới. Những giống lúa có bông trổ thoát hoàn toàn thường cho tỷ lệ hạt chắc cao hơn [6].

1.3. QTL và lập bản QTL

QTL (Quantitative Trait Locus) hay còn gọi là lô-cut tính trạng số lượng hoặc lô-cut đặc điểm định lượng dùng để chỉ một hoặc nhiều lô-cut gen quy định tính trạng số lượng. Loại tính trạng số lượng (Quantitative trait) thường phụ thuộc vào hoạt động tương tác của nhiều gen với nhau và với môi trường. Do đó, kiểu hình những tính trạng này thường khác nhau giữa các cá thể cùng loài, luôn thay đổi trị số trong một phạm vi nhất định, nên tạo ra sự phân bố các giá trị kiểu hình liên tục trong quần thể. Tính trạng số lượng thường là tính trạng đa gen (polygenic trait), quá trình di truyền các gen này và biểu hiện của chúng phức tạp. Vì một tính trạng số lượng thường là tính trạng đa gen, nên một QTL thường bao gồm nhiều gen và chiếm một khu vực nhất định trong bộ gen (genom) chịu trách nhiệm quy định sự hình thành cũng như biến đổi của tính trạng số lượng tương ứng. Như vậy, một QTL là một vùng nhiễm sắc thể cụ thể hoặc các lô-cut gen chứa các trình tự DNA đặc trưng liên quan đến tính trạng số lượng mà nó điều khiển. Một khi đã xác định được QTL, nhà nghiên cứu hoàn toàn có khả năng xác định gen hoặc các gen cụ thể trong QTL đó, dùng kỹ thuật đánh dấu phân tử theo dõi. Phương pháp này gọi là lập bản đồ QTL (QTL mapping) và giải trình tự các gen gây ra sự biến đổi tính trạng [135].

Lập bản đồ tính trạng số lượng (QTL mapping) trên toàn bộ hệ gen suy ra mối quan hệ giữa kiểu gen tại các vị trí gen khác nhau và kiểu hình cho một tập hợp các tính trạng số lượng về số lượng, vị trí gen, sự ảnh hưởng và tương tác của QTL. Mục đích cơ bản của lập bản đồ QTL là xác định vị trí các vùng nhiễm sắc thể có ảnh hưởng đáng kể đến sự biến đổi của các tính trạng số lượng trong quần thể. Việc xác định vị trí này rất quan trọng đối với việc nhận

dạng các gen cuối cùng có trách nhiệm và hiểu được các cơ chế di truyền của sự biến đổi trong quần thể. Lập bản đồ QTL giúp chúng ta hiểu có bao nhiêu QTL đóng góp đáng kể vào sự biến đổi tính trạng trong quần thể, bao nhiêu biến đổi là do tác động cộng gộp của QTL và bao nhiêu là do tác động chi phối và tác động lấn át gen của QTL. Lập bản đồ QTL cũng cho biết bản chất của mối tương quan di truyền giữa các tính trạng khác nhau trong một vùng gen, biểu hiện đa hình hay liên kết gần là gì, QTL có tương tác với các môi trường không. Những câu hỏi này liên quan đến cấu trúc di truyền của các tính trạng số lượng trong quần thể và có liên quan mật thiết đến nhiều ứng dụng trong di truyền số lượng, chẳng hạn như dự đoán hoặc chọn lọc có sự hỗ trợ của chỉ thị di truyền và độ thâm nhập của gen có sự hỗ trợ của dấu hiệu đánh dấu [122].

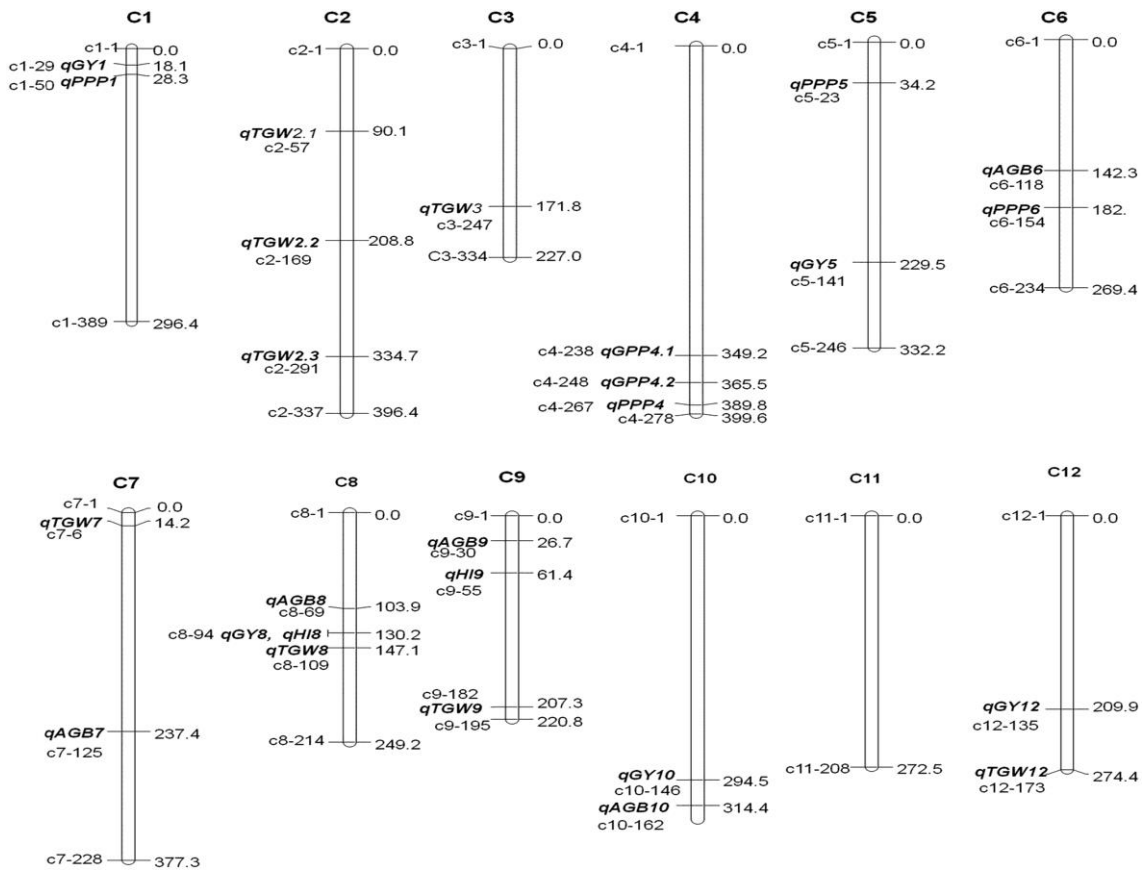
Dữ liệu đánh dấu có tính phân loại và có thể được phân loại thành các danh mục khác nhau và được ghi lại ở dạng kỹ thuật số, chẳng hạn như 1 hoặc 0 đối với sự có mặt hoặc không có dải phân tử cụ thể tại một hoặc hai kiểu gen đánh dấu (đồng hợp tử và dị hợp tử) cho phép lai chéo quần thể từ hai dòng cận huyết. Dựa trên phân tích phân li, các dấu hiệu này có thể được sắp xếp theo thứ tự trong các nhóm liên kết hoặc tuyến tính trên nhiễm sắc thể để biểu diễn bản đồ liên kết di truyền [122]. Trong khi dữ liệu đánh dấu chứa thông tin về sự phân ly của bộ gen trong quần thể, thì dữ liệu về tính trạng số lượng chứa thông tin về sự biến đổi của các tính trạng trong quần thể. Hai tập dữ liệu được kết nối bởi QTL. Một phần của sự biến đổi tính trạng là do sự phân li của QTL liên kết với một số chỉ thị trong hệ gen. Vì vậy, nhiệm vụ thống kê của lập bản đồ QTL là liên hệ sự biến đổi tính trạng số lượng với sự biến đổi chỉ thị di truyền bao gồm số lượng, vị trí, sự liên kết và tương tác của các gen ảnh hưởng đến các tính trạng số lượng cần quan tâm [56].

Ở cây trồng, đặc biệt là ở lúa, các tính trạng nông học quan trọng như năng suất thường là các tính trạng số lượng chịu ảnh hưởng của nhiều gen khác nhau và chịu tác động của môi trường. Việc xác định các lô-cut quy định tính trạng số lượng (QTL) dựa trên phân tích kiểu gen và kiểu hình ở các quần thể phân li và áp dụng các công cụ thống kê nhất định. Phương pháp nghiên cứu các tính trạng số lượng là xác định vị trí các lô-cut gen quy định tính trạng số lượng dựa trên sự liên kết của chúng với các chỉ thị phân tử [46].

1.4. Các QTL liên kết với tính trạng năng suất và các yếu tố cấu thành năng suất ở lúa

Tính trạng cấu thành năng suất là một tính trạng nông học phức hợp do nhiều lô-cut tính trạng số lượng quy định (QTL), thường kiểu hình biểu hiện liên tục trong các quần thể phân li từ các tổ hợp lai của các dòng lúa thuần. Đa phần các QTL quy định tính trạng năng suất biểu hiện ảnh hưởng di truyền không đáng kể và rất khó xác định. Các QTL này đóng vai trò quan trọng trong việc kiểm soát tính trạng năng suất và đã được ứng dụng rộng rãi trong các giống lúa thương mại, do vậy lập bản đồ chi tiết và nhân bản các QTL quan trọng sẽ mang lại nhiều lợi ích trong công tác chọn giống. Ba tính trạng quan trọng quy định yếu tố cấu thành năng suất bao gồm: Số bông trên khóm, số hạt trên bông, và khối lượng nghìn hạt. Trong đó các tính trạng này lại phụ thuộc vào khả năng đẻ nhánh, số nhánh hữu hiệu và khả năng phát triển của hạt, như hạt chắc, hạt lép.... Phân tích cơ sở di truyền của các tính trạng này bằng việc lập bản đồ để chọn tạo ra các giống lúa cho năng suất cao hiện đã và đang được nghiên cứu ở nhiều quốc gia trên thế giới. Trong nhiều thập kỷ, bằng cách xây dựng bản đồ QTL, nhiều QTL/gen quy định tính trạng cấu thành năng suất đã được xác định trên rất nhiều quần thể từ các tổ hợp lai giữa giống hay các loài phụ. QTL/gen liên kết với tính trạng năng suất và yếu tố cấu thành năng suất được xác định trên tất cả các nhiễm sắc thể của lúa [4].

Kết quả nghiên cứu của Zhu và các cộng sự (2017) [129] đã tiến hành phân tích QTL cho sáu tính trạng năng suất bao gồm số bông trên cây, số hạt trên bông, năng suất hạt trên cây, khối lượng nghìn hạt, sinh khối trên mặt đất và chỉ số thu hoạch bằng cách sử dụng các chỉ thị SNP trong quần thể các dòng lai tái tổ hợp (RILs) giữa giống *Japonica* nhiệt đới Francis thuần chủng và giống *Indica Guanghui 998* (R998). Tổng cộng 26 QTL đã được phát hiện bằng cách sử dụng bản đồ di truyền mật độ cao bao gồm 3016 mã đánh dấu. Mười chín trong số 26 QTL từ R998 có ảnh hưởng có lợi đến các tính trạng năng suất (hình 1.1)



Hình 1.1. Bản đồ vị trí các QTL liên kết với tính trạng năng suất trên 12 nhiễm sắc thể ở lúa

(PPP - số bông/cây; GPP - số hạt/bông; GY - năng suất hạt trên mỗi cây; TGW - khối lượng nghìn hạt; AGB - sinh khối trên mặt đất; HI - chỉ số thu hoạch; C - nhiễm sắc thể). Nguồn: Zhu và cộng sự (2017) [129]

Một nghiên cứu gần đây của nhóm nghiên cứu Donde và các cộng sự (2020) [39] đã sử dụng 85 chỉ thị phân tử SSR để xác định các QTL có liên quan đến năng suất lúa và các tính trạng liên quan của tập đoàn 60 giống lúa NPT (New Plant Types). Kết quả nghiên cứu đã chỉ ra có 30 QTL mới liên kết với 16 chỉ thị phân tử SSR được xác định có liên quan đến 11 chỉ tiêu: số nhánh/cây (*qTL-6.1, qTL-11.1, qTL-4.1*), chiều dài bông (*qPL-1.1, qPL-5.1, qPL-7.1, qPL-8.1*), chiều dài lá đòng (*qFLL-8.1, qFLL-9.1*), chiều rộng lá đòng (*qFLW-6.2, qFLW-5.1, qFLW-8.1, qFLW-7.1*), tổng số hạt (*qTG-2.2, qTG-a7.1*), khối lượng nghìn hạt (*qTGW-a1.1, qTGW-a9.2, qTGW-5.1, qTGW-8.1*), số hạt chắc (*qFG-7.1*), tỷ lệ chiều dài - chiều rộng hạt (*qSlb - 3.1*), chiều cao cây (*qPHT-6.1, qPHT-9.1*), số ngày ra hoa đến 50% (*qFD-1.1*) và năng suất hạt mỗi ngày (*qYLD-5.1, qYLD-6.1a, qYLD-11.1*).

Trong những năm qua, nhiều QTL liên kết với các tính trạng năng suất và các yếu tố cấu thành năng suất đã được xác định trên 12 nhiễm sắc thể ở lúa bằng phương pháp lập bản đồ QTL nhưng chưa có QTL nào được tìm thấy liên kết với cả hai tính trạng số gié cấp hai/bông và số hạt/bông.

1.5. Các quần thể lập bản đồ QTL liên quan đến tính trạng năng suất lúa

Lập bản đồ QTL của một tính trạng mục tiêu là xác định vị trí QTL đó trên nhiễm sắc thể và phân tích đặc tính di truyền cho tính trạng quan tâm thông qua mối liên kết giữa chỉ thị di truyền và biến dị kiểu hình. Để lập bản đồ thì việc phát triển quần thể và xây dựng bản đồ liên kết, đánh giá kiểu hình là việc làm cần thiết cho nghiên cứu phân tích QTL [4].

Ở thực vật, có hai phương pháp chính: 1) sử dụng các quần thể tạo ra từ các phép lai được kiểm soát (F_1 tự thụ phấn hoặc lai trở lại thế hệ bố mẹ.); 2) sử dụng các quần thể tự nhiên, giao phấn tự do không kiểm soát. Đối với cây lúa, hầu hết đều sử dụng phương pháp 1) để nghiên cứu. Phương pháp này cần sử dụng một bộ chỉ thị phân tử trên toàn bộ hệ gen (SNP - Đa hình đơn

nucleotit; RFLP - đa hình chiều dài đoạn cắt giới hạn, AFLP - đa hình đoạn cắt khuếch đại...) và lập bản đồ liên kết các chỉ thị đó trên quần thể đang phân tích. Sau đó sử dụng các công cụ tính toán để có thể lập nhóm liên kết và tính khoảng cách di truyền giữa các chỉ thị, xác định giá trị kiểu hình tính trạng nghiên cứu trên từng cá thể. Bằng các chương trình thống kê, tiến hành khoanh vùng các vị trí có thể có liên quan đến tính trạng cần nghiên cứu, chọn ra vùng QTL tiềm năng nhất để phân lập, sau đó cô lập đến vị trí từng gen [134].

Các quần thể lập bản đồ QTL dựa trên sự phát triển của các dòng lai F_1 có nguồn gốc từ các cặp lai chéo giữa các giống khác nhau, thường có sự khác biệt đối lập đối với tính trạng quan tâm, ví dụ như giữa các giống có cấu trúc bông to và các giống có cấu trúc bông nhỏ. Từ đó, một quần thể phân li với sự mất cân bằng do di truyền liên kết (LD) lớn giữa các lô-cut được thành lập, là nền tảng của phân tích liên kết di truyền. Quần thể lập bản đồ này có thể được lai chéo hoặc tự thụ phấn do đó làm giảm LD thông qua tái tổ hợp. Kết quả là chỉ các chỉ thị được liên kết chặt chẽ với QTL sẽ cho phép phát hiện mối liên kết và do đó độ phân giải bản đồ được tăng lên. Nhiều loại quần thể lập bản đồ có thể được sử dụng cho phân tích QTL, chẳng hạn như quần thể F_2 [25], quần thể đơn bội (Doubled haploids - DHs) [21], [46], [117], quần thể của các dòng thuần chủng tái tổ hợp (Recombinant Inbred Lines-RILs) [18], [76], quần thể lai trở lại cận giao tái tổ hợp (Backcross Recombinant Inbred Lines – BILs) [87], [107].

Quần thể F_2 có nguồn gốc từ cặp lai giữa 2 giống bố mẹ và sau đó tự thụ phấn chứa đầy đủ thông tin di truyền của 2 giống bố mẹ. Về lý thuyết nó có thể được sử dụng để xác định QTLs với các hiệu ứng phụ và trội trong toàn bộ bộ gen của cây lúa. Vì vậy nó được sử dụng rộng rãi để phân tích cơ sở di truyền của năng suất lúa và các yếu tố cấu thành năng suất [25]. Tuy nhiên trong quần thể F_2 , mỗi cá thể có một kiểu gen riêng biệt, dẫn đến các giá trị

đặc điểm không đáng tin cậy gây ra bởi các hiệu ứng môi trường chứ không phải hiệu ứng di truyền. Quần thể F_2 đã được sử dụng trong phát hiện các QTL liên quan đến năng suất lúa và các yếu tố cấu thành năng suất [130]; QTL liên quan đến chiều dài hạt [57]; trong lập bản đồ QTL liên quan đến chiều dài bông và mật độ hoa/bông ở lúa [25].

Các quần thể đơn bội (DHs) được phát triển bằng cách nuôi cấy bao phấn từ một cây F_1 có nguồn gốc từ hai bố mẹ. Một nghiên cứu gần đây của Xu và các cộng sự (2020) [117] sử dụng một quần thể đơn bội kép (DHs) có nguồn gốc từ phép lai giữa cây *Japonica* “Maybelle” và *Indica* “Baiyeqiu” để phân tích QTL liên quan đến cấu trúc bông lúa cho thấy tổng số 24 QTL liên quan đến cấu trúc bông và hình dạng hạt đã được phát hiện trên nhiễm sắc thể 1, 2, 3, 4, 5, 6, 7, 9, 10 và 12, và ba QTL tương tác tñnh điện đã được phát hiện. Năm QTL cho chiều dài bông được phát hiện trên nhiễm sắc thể 2, 3, 4 và 7, với điểm LOD dao động từ 2,75 đến 6,13 đã giải thích 4,74 - 15,02% sự biến đổi kiểu hình. Bốn QTL làm tăng chiều dài bông *qPL2*, *qPL3*, *qPL4-1* và *qPL7*, thể hiện các hiệu ứng phụ âm và các alen từ “Baiyeqiu”. Đối với số lượng gié cấp hai (SBN), bốn QTL là *qSBN1*, *qSBN3*, *qSBN4* và *qSBN7* lần lượt nằm trên các nhiễm sắc thể số 1, 3, 4 và 7, giải thích được 5,83 - 11,47% sự biến đổi kiểu hình và có điểm LOD dao động từ 3,04 đến 5,12. Ba trong số các QTL hiển thị các hiệu ứng phụ âm, cho thấy rằng các alen từ “Baiyeqiu” làm tăng SBN.

Do hiệu quả của nuôi cấy bao phấn phụ thuộc rất nhiều vào nền di truyền nên quần thể DHs bị giới hạn trong nghiên cứu lập bản đồ QTL. Trong khi đó, quần thể của các dòng thuần chủng tái tổ hợp (Recombinant Inbred Lines - RILs) là lựa chọn chính cho cây tự thụ phấn như lúa [126]. Do những lợi thế thu được từ sự đồng hợp tử và sự tái tổ hợp có hiệu quả của chúng. RILs được phát triển từ một hạt F_2 duy nhất, từ mỗi cây F_2 một hạt giống duy

nhất được thu hoạch và trồng thành cây F_3 , từ đó một hạt giống duy nhất được thu hoạch và trồng thành cây F_4 , v.v. Trong quá trình này, mức độ dị hợp tử trên mỗi lô-cut được giảm đi một nửa trong mỗi thế hệ, vì vậy quần thể trở nên đồng hợp tử ở thế hệ F_8 trở đi. Vì lý do này, RILs được coi là các quần thể đồng hợp tử đồng nhất có thể được nhân giống vô thời hạn và chỉ cần phải phân tích kiểu gen để xác định các chỉ thị phân tử chỉ một lần. Ngoài ra, trong quá trình lai, thực vật trải qua một vài vòng giảm phân, tăng tần suất tái tổ hợp giữa các chỉ thị liên kết chặt chẽ. Các quần thể RILs có thể được sử dụng để lập bản đồ QTL rất hiệu quả bởi vì ảnh hưởng của môi trường đến đặc điểm định lượng có thể giảm đi nhiều bằng cách đánh giá một số cây có cùng kiểu gen thay vì chỉ một cây duy nhất. Một khi đã được phân tích kiểu gen, quần thể RILs có thể được sử dụng cho các nghiên cứu không giới hạn các tính trạng khác nhau trong các môi trường khác nhau để loại bỏ hiệu ứng môi trường cũng như sai sót thực nghiệm, cho phép lập bản đồ QTL so sánh hiệu quả. Hơn nữa việc ứng dụng các chỉ thị phân tử cho phép chọn lọc các dòng đồng hợp tử ngay ở thế hệ F_2 , rút ngắn thời gian thu được quần thể RILs đồng hợp tử xuống còn 3 thế hệ thay vì 8 - 10 thế hệ.

Nghiên cứu xác định các lô-cut tính trạng số lượng (QTLs) chi phối năng suất và các tính trạng liên quan của nó bằng cách sử dụng quần thể lai tái tổ hợp (RILs) có nguồn gốc từ giống lúa lai phổ biến, KRH-2 (IR58025A/KMR3R). Một bản đồ di truyền kéo dài 294,2 cM đã được xây dựng với 126 lô-cut lặp lại trình tự đơn giản (SSR) phân bố đồng đều trên bộ gen lúa. Phân tích QTL bằng cách sử dụng thông tin về kiểu hình và kiểu gen đã xác định được tổng số 22 QTL. Trong đó, năm QTL ảnh hưởng chính được xác định đối với các tính trạng sau: tổng năng suất hạt/cây (*qYLD3-1*), khối lượng bông (*qPW3-1*), chiều cao cây (*qPH12-1*), chiều rộng lá cờ (*qFLW4-1*) và chiều dài bông (*qPL3-1*), giải thích 20,23 - 22,76% phương sai kiểu hình

với điểm LOD nằm trong khoảng 6,5 - 10,59. Một số vùng gen kiểm soát một số tính trạng như tổng năng suất hạt/cây (*qYLD3-1*) và chiều dài bông (*qPL3-1*) được xác định trên nhiễm sắc thể số 3 [60].

Quần thể RIL có nguồn gốc từ cặp lai chéo giữa 2 giống *Indica* Minghui 63 and Zhenshan 97 đã được sử dụng rộng rãi để lập bản đồ QTL tính trạng cấu trúc bông [25], [71], [87], [128]. Các QTL liên quan đến hình dạng hạt (*GS3*), số hạt/bông, ngày ra hoa và chiều cao cây (*Ghd7*) đã được phân lập thành công [41], [118].

Sử dụng quần thể thể BIL (Backcross inbred lines – lai trở lại cận giao tái tổ hợp) để phân tích QTL, Ashikari và cộng sự (2005) [14] đã xác định được năm QTL liên quan đến số lượng hạt/bông, trong đó có QTL *Gn1* làm tăng số lượng hạt tới 92 hạt; Sun và cộng sự (2017) [101] đã phát hiện 12 QTL liên quan đến số hạt/bông, chiều dài bông, số gié cấp một và số gié cấp hai đều nằm trên nhiễm sắc thể số 6, trong đó QTL *L6-5* xác định chiều dài trục bông. Hai QTL đã được phát hiện trên nhánh dài của nhiễm sắc thể 12: *qTSN12.1* trong quần thể BC₄F₂ của YTH63/IR64 và *qTSN12.2* trong quần thể BC₄F₃ của YTH83/IR 64 được xác định có chức năng quan trọng trong việc thúc đẩy quá trình phân hóa các gié trên bông lúa ở giai đoạn hình thành tế bào [95]. Các QTL liên quan đến chiều dài bông nằm trên nhiễm sắc thể số 4, 9, 10, trong đó có QTL *PL9* là QTL chính liên quan đến tăng chiều dài bông lúa *O.minuta* [128]. QTL xác định chiều dài hạt là *qPL2* nằm ở vị trí 85,01 cM và hai QTL xác định mật độ hoa/bông *qSD3.1* và *qSD3.2*, lần lượt nằm ở 28,91 và 39,51 cM cả ba QTL này đều nằm trên nhiễm sắc thể số 3 [25].

Bản đồ cơ sở không thể xác định được vị trí chính xác của một QTL, do vậy các thí nghiệm tiếp theo cần phải được tiến hành để xác định được chức năng sinh học của các QTL mục tiêu. Phát triển dòng cận đẳng gen (NIL) là một chiến lược hiệu quả để kiểm tra QTL [95], [101]. Dòng NIL bao gồm các

vùng QTL mục tiêu phân li trong một nền di truyền dạng đồng hợp. Nói chung, dòng NIL được phát triển bằng thực hiện lai trở lại liên tục với dòng mẹ hồi quy bằng chọn lọc nhờ chỉ thị phân tử (MAS). Trong đó, chỉ chọn lọc các cá thể mang QTL mục tiêu của nền di truyền của dòng mẹ, sau đó để phát triển dòng NIL. Nhiều QTL kiểm soát các tính trạng năng suất đã được kiểm tra bằng phương pháp cổ điển này. Tuy nhiên, lai trở lại kết hợp với chọn lọc nhờ chỉ thị phân tử (MAS) sẽ giảm được thời gian và công sức sàng lọc [4].

Dòng lai nhập (introgression line - IL) và dòng thay thế phân đoạn nhiễm sắc được phát triển bằng việc thực hiện lai trở lại liên tục với dòng mẹ nhận gen, cũng có thể được ứng dụng để kiểm tra QTL, lập bản đồ chi tiết và chọn tạo các giống lúa siêu năng suất [90].

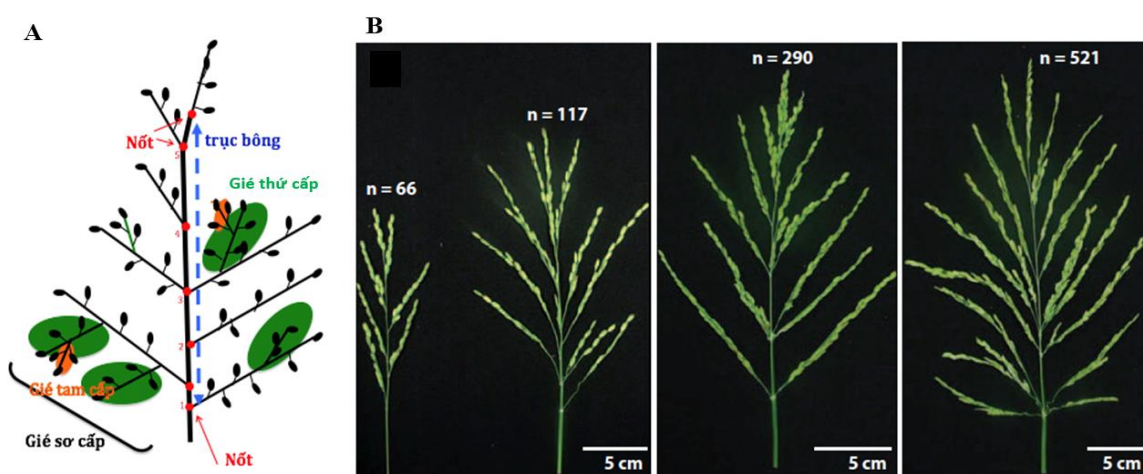
Các quần thể lập bản đồ có thể được đánh giá kiểu gen với bất kỳ chỉ thị phân tử nào, chẳng hạn như các chỉ thị dựa vào kỹ thuật PCR như: SSR (Simple Sequence Repeat), CAPS (Cleaved Amplified Polymorphic Sequences), hoặc các chỉ thị được xây dựng từ các dữ liệu giải trình tự như: SNPs (Single Nucleotide Polymorphism) hoặc INDELs (Inserts/Deletions).

1.6. Cấu trúc bông lúa và các QTL/gen liên quan đến cấu trúc bông lúa

1.6.1. Cấu trúc bông lúa

Cấu trúc bông là một trong những đặc điểm nông học quan trọng liên quan đến năng suất lúa [115]. Cấu trúc bông lúa thể hiện qua kích thước, số lượng và hình dáng của bông lúa. Theo đó, cả sự phân nhánh và hình thành bông lúa đều có thể ảnh hưởng đến cấu trúc bông [121]. Các chỉ số ước tính cấu trúc bông lúa thường bao gồm số hạt/bông (SpN), chiều dài bông (PL), số lượng gié cấp một (PBN) và gié cấp hai (SBN) [101]. Ngoài ra còn một số đặc điểm liên quan đến cấu trúc bông bao gồm hình dạng hạt, có thể được suy ra bởi chiều dài hạt (GL), chiều rộng hạt (GW), và tỷ lệ chiều dài - chiều rộng hạt (LWR) [17].

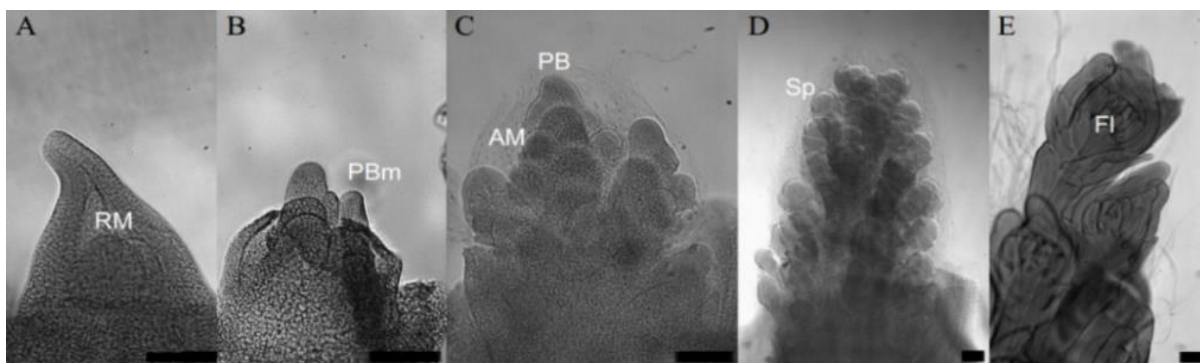
Cấu trúc bông lúa là một trong những tính trạng hình thái quan trọng quyết định tiềm năng năng suất, được chọn lọc qua quá trình thuần hóa. Bông lúa là một cấu trúc dạng nhánh, bao gồm một trục chính, các gié cấp một mọc ra từ trục chính, sau đó là các gié bậc cao hơn (gié cấp hai, cấp ba, cấp bốn) và cuối cùng là hoa lúa, mọc trực tiếp trên các gié (hình 1.2A). Số lượng hoa lúa quyết định số lượng hạt trên bông, trực tiếp ảnh hưởng đến năng suất cây trồng. Quá trình phân nhánh bông phụ thuộc vào hoạt động của các tế bào gốc chồi nách trong suốt giai đoạn hình thành và phát triển sớm của bông. Cấu trúc bông lúa rất đa dạng, không chỉ trong cùng loài (giữa giống thuần và giống hoang dại) mà còn giữa các loài khác nhau (lúa Châu Á và lúa Châu Phi). Cấu trúc bông lúa ở các loài thuần chủng thay đổi từ dạng bông nhỏ, chỉ có một vài gié cấp một và gié cấp hai, mang ít hạt đến dạng bông to, phân nhánh nhiều và mang nhiều hạt hơn so với các loài tổ tiên (Hình 1.2B) [104], [120].



Hình 1.2. Cấu trúc bông lúa (A) và sự thay đổi qua quá trình thuần hóa (B),
n: Số hạt/bông. Nguồn: Yoshida, 2011 [120]

Bông lúa là kết quả quá trình hoạt động của các mô phân sinh, chúng được hình thành và biệt hóa liên tục trong quá trình phát triển của cây [63]. Sự hình thành bông lúa non được chia thành hai giai đoạn: giai đoạn sớm là sự hình thành và kéo dài của mô phân sinh trục (Hình 1.3A), hình thành các

mô phân sinh gié cấp một (Hình 1.3B), kéo dài các mô phân sinh gié cấp một, hình thành các mô phân sinh gié cấp cao hơn (Hình 1.3C). Giai đoạn muộn là sự hình thành mô phân sinh hoa (Hình 1.3D), sự biệt hóa tạo thành các bộ phận của hoa và sự phát triển của chúng cũng được quan sát ở giai đoạn này (Hình 1.3E).



Hình 1.3. Các giai đoạn hình thành bông lúa non quan sát dưới kính hiển vi
A : Giai đoạn phát sinh mô phân sinh trục bông, B : Hình thành và kéo dài mô phân sinh gié cấp một, C : Kéo dài gié cấp một và hình thành gié cấp hai, D : Hình thành hoa, E : Sự biệt hóa các cơ quan sinh sản của hoa. RM : mô phân sinh trục ; PBm : mô phân sinh gié cấp một; AM : mô phân sinh gié cấp hai; PB : Gié cấp một; Sp : Cụm hoa ; Fl : hoa. Nguồn: Tanaka, 2014 [104]

1.6.2. Các QTL/gen liên quan đến cấu trúc bông lúa

Nâng cao năng suất cây trồng là một trong những ưu tiên hàng đầu trong chương trình chọn tạo giống. Trong đó, cải thiện cấu trúc bông lúa đã được chứng minh là một chiến lược thành công [71]. Cấu trúc bông lúa được quyết định bởi chiều dài trục bông (PL), số gié cấp một (PBN), số gié cấp hai (SBN), số lượng hạt (SpN) [67]. Các tính trạng này đều được kiểm soát bởi các lô-cut tính trạng số lượng (QTLs) [102].

Các QTL quy định các tính trạng cấu trúc bông như số gié cấp một, số gié cấp hai, số gié cấp ba cũng đã được lập bản đồ và nghiên cứu [14], [124]. Bốn tính trạng bông: số hạt, chiều dài bông, số gié cấp một và số gié cấp hai được đánh giá ảnh hưởng của QTL bằng cách sử dụng phương pháp lập bản

đồ khoảng cách tổng hợp (ICIM) trong các quần thể lai trở lại (BIL152, BIL196a, BIL196b và BIL196b-156) đã xác định có 12 QTL nằm trên nhiễm sắc thể số 6. Các QTL này được phát hiện có liên kết chặt chẽ với ba khoảng nhiễm sắc thể RM7213 đến RM19962, RM20000 đến RM20210 và RM412 đến RM20595. Trong đó có 1 QTL (*PL6-5*) quy định tính trạng chiều dài bông đều có ở cả bốn quần thể nằm trong vùng vùng 1,3Mb trên nhiễm sắc thể 6 [101].

Số hạt/bông là tính trạng nông học quan trọng để tăng năng suất lúa, được quy định bởi sự hình thành bông [6]. Số hạt/bông thường tỷ lệ thuận với số gié/bông. Trong suốt hơn hai thập kỷ qua, nhiều QTL/gen kiểm soát tính trạng cấu trúc bông lúa đã được xác định và phân tích. Một số QTL/gen liên kết đến sự hình thành phát triển gié đã được xác định bằng phân tích đột biến [116].

Nghiên cứu về tính trạng số gié/bông, Komatsu và cộng sự năm 2003 đã phát hiện ra vai trò của gen *LAX1* mã hóa một chuỗi xoắn (bHLH) là yếu tố phiên mã rất cần cho quá trình duy trì mô phân sinh ở nhánh hoa. *LAX1* liên quan tới tất cả các kiểu phân chia chồi trong suốt quá trình phát triển của lúa, đột biến gen *LAX1* làm giảm số gié cấp một và gié cấp hai/bông. Gen *LAX1* đã được các tác giả nhân dòng thành công bằng chỉ thị phân tử CAPS. Cũng trong nghiên cứu này, các tác giả còn phát hiện ra gen *SMALL PANICLE (SPA)*. Gen này có vai trò điều chỉnh sự hình thành chồi bên, khi xảy ra đột biến *SPA* thì số gié cấp một và gié cấp hai của lúa đều bị giảm [61].

Năm 2005, Ashikari và cộng sự đã xác định được một QTL làm tăng số hạt/bông là *GRAIN NUMBER1 (Gn1a)*. Trong nghiên cứu này tác giả đã sử dụng một dòng lúa *Indica* là Habataki có số hạt/bông là 300 hạt và một dòng lúa *Japonica* là Koshihikari có số hạt/bông là 160 hạt. Nghiên cứu đã xác định được 5 QTL quy định tính trạng tăng số hạt/bông, trong đó *Gn1a* định vị trên vai ngấn của nhiễm sắc thể số 1 dẫn tới sự khác biệt 44% số hạt/bông của hai dòng Habataki và Koshihikari [14].

Kích thước bông chủ yếu được xác định bởi số lượng và độ dài của các gié cấp một và cấp hai [6, tr.76]. Nghiên cứu về cơ chế điều khiển sự phát triển các gié lúa, Li và cộng sự năm 2009 [66] đã nhân dòng thành công gen *SHORT PANICLE (SP1)* dựa trên phương pháp lập bản đồ. *SP1* gây ra kiểu hình trùn ngắn, mã hóa một protein vận chuyển giả định tham gia vào quá trình dài bông ở lúa.

Cấu trúc bông chủ yếu được xác định bởi sự sắp xếp của các gié cấp một, gié cấp hai và mật độ hạt. Bông thẳng đứng là một tính trạng nông học quan trọng liên quan chặt chẽ với năng suất hạt. Gen *DEP1* mã hóa protein PEBP kiểm soát số gié, mật độ hạt và kiểu bông thẳng đứng. *DEP1* biểu hiện ở kiểu hình là tăng gié cấp một, cấp hai và số hạt/bông. Trong khi đó, *DEP2* mã hóa protein cụ thể và biểu hiện đáng kể ở bông non, kiểm soát sự phát triển và độ dài bông. Đột biến gen *DEP2* biểu hiện kiểu hình ở mật độ và bông thẳng [50].

Nghiên cứu về tính trạng tăng số hạt/bông, Linh và cộng sự (2008) đã lập bản đồ QTL *yd7* liên kết tính trạng tăng năng suất từ lúa hoang *O.minuta* trên nhiễm sắc thể số 7. Sự có mặt của *yd7* trên giống lúa trồng đại trà Hwaseongbyeon làm tăng 15 - 20% năng suất lúa [70].

Nghiên cứu lập bản đồ QTL liên quan đến số hạt/bông ở các dòng nhập nội có số hạt/bông cao có nguồn gốc từ phép lai giữa giống *Indica* IR64 và giống NPT (New Plant Type) mới đây của tác giả Sasaki 2017 đã xác định được 2 QTL nằm trên nhánh dài nhiễm sắc thể số 12 bao gồm: QTL *qTSN12.1* trong quần thể BC₄F₂ của YTH63/IR 64 và QTL *qTSN12.2* trong quần thể BC₄F₃ của YTH83/IR 64. Nghiên cứu cũng chỉ ra chức năng quan trọng của hai QTL này trong việc thúc đẩy sự phân nhánh của chồi ở giai đoạn hình thành bông lúa. Phân tích kiểu hình chi tiết về các tính trạng liên quan đến năng suất của quần thể IR 64-NIL12 mang *qTSN12.2* cho thấy dòng

này có số bông thấp hơn, thân dài hơn, lá dài, rộng hơn so với IR 64. Mặt khác, khối lượng một nghìn hạt có xu hướng cao hơn và năng suất hạt trên mỗi mét vuông ở IR 64-NIL12 cũng lớn hơn ở IR 64. Các QTL mới được xác định có ý nghĩa cho việc cải thiện di truyền về tiềm năng năng suất của các giống thuộc nhóm *Indica* [95].

Như vậy, trong nhiều năm qua đã có nhiều nghiên cứu về các QTL/gen quy định tính trạng liên quan bông lúa bởi vì đây là yếu tố quan trọng cấu thành năng suất lúa.

Trong số các QTL đã được công bố liên quan đến cấu trúc bông chưa có QTL nào được phát hiện liên quan đến cả hai tính trạng số gié cấp hai/bông và số hạt/bông. Đây là hai tính trạng quan trọng quyết định tiềm năng năng suất lúa.

1.7. Chỉ thị phân tử CAPS và ứng dụng trong nghiên cứu lập bản đồ QTL liên quan đến tính trạng năng suất lúa

Phần lớn các chỉ thị phân tử sử dụng trong di truyền chọn giống dựa vào kỹ thuật PCR và sự lặp lại ngẫu nhiên của một số nucleotit trong hệ genom sinh vật nên tính đặc hiệu không cao. Cùng với sự phát triển của công nghệ giải trình tự, nhất là sự ra đời của các công nghệ giải trình tự thế hệ mới đã cho phép tiến hành hàng loạt dự án giải trình tự giống lúa khác nhau trên thế giới. Từ việc phân tích kết quả giải trình tự, hàng triệu điểm đột biến đa hình đơn nucleotide (SNPs) đã được phát hiện và được lập bản đồ phục vụ cho các nghiên cứu trên toàn hệ gen. Một trong những tính ứng dụng thực tiễn của SNP là dùng để xây dựng chỉ thị CAPS nhằm phục vụ cho việc phân biệt các giống khác nhau. Nguyên tắc của chỉ thị phân tử CAPS dựa trên sự thay đổi của một nucleotit (SNP) trong vị trí cắt của enzyme giới hạn. Từ đó các mẫu sẽ được thiết kế nhằm nhân đoạn DNA chứa SNP, sau đó chạy phản ứng PCR với mồi CAPS và cắt sản phẩm PCR bằng enzyme giới hạn thành hai

đoạn có chiều dài khác nhau. Những giống mà vị trí cắt của enzyme giới hạn không bị đột biến thì sản phẩm PCR sẽ bị cắt thành hai mảnh. Ngược lại đối với giống mang SNP trong vùng cắt của enzyme giới hạn thì enzyme không nhận biết được vị trí cắt đặc thù của nó dẫn đến sản phẩm PCR không bị cắt. Nhờ vào kỹ thuật điện di trên gel agarose sẽ dễ dàng phân biệt được các giống khác nhau dựa vào đa hình chiều dài của sản phẩm cắt bằng enzyme giới hạn.

Chỉ thị phân tử CAPS [62] đại diện cho một nhóm chỉ thị đặc hiệu đã được ứng dụng thành công, đặc biệt là trong lĩnh vực sinh học thực vật. Nguyên tắc của CAPS rất đơn giản và dựa trên ba bước: (1) PCR với môi đặc hiệu; (2) cắt sản phẩm PCR với enzyme giới hạn; và (3) điện di sản phẩm cắt trên gel agarose. Trên thực tế, chỉ thị phân tử CAPS kết hợp sử dụng kỹ thuật PCR với các phương pháp cổ điển RFLP, chỉ khác là nó dựa trên sự khuếch đại của các đoạn DNA ngắn thay vì toàn bộ bộ gen [45], [49], [73]. Thiết kế môi cho chỉ thị CAPS có thể dựa trên trình tự nucleotit từ các cơ sở dữ liệu sẵn có hoặc giải trình tự vùng cần khuếch đại [49], [53], [71], [109].

Bước đầu tiên của CAPS thực ra là một PCR thông thường. Ứng dụng enzyme cắt giới hạn là bước thứ hai và ở đây xác suất đa hình di truyền giữa các mẫu DNA đóng vai trò trung tâm. Khác biệt di truyền trong DNA chủ yếu ở dạng đa hình đơn nucleotit (SNP), hoặc chèn xóa (InDel) [44], [54], [59], [115]. Những thay đổi di truyền trong các vị trí nhận biết của các enzyme giới hạn là những điểm quan trọng để phát triển và ứng dụng của các chỉ thị CAPS. Một đoạn khuếch đại với một vị trí nhận diện chính xác của một enzyme giới hạn sẽ bị cắt thành hai phân đoạn bằng enzyme giới hạn. Ngược lại, chỉ có một đoạn duy nhất có thể được quan sát thấy khi vị trí cắt của enzyme giới hạn đã bị đột biến và thay đổi khiến cho đoạn khuếch đại không bị cắt.

Để xác định vị trí chính xác của SNP trong các vị trí nhận biết của các enzyme giới hạn, sẽ rất hữu ích khi có trình tự của các đoạn khuếch đại [28], [72], [73], [77]. Giải trình tự thế hệ mới (Next Generation Sequencing - NGS) là một phương pháp giải trình tự hiện đại và rất phổ biến đã giúp xác định hàng triệu SNPs cho hàng ngàn giống lúa khác nhau (Rice SNP - Seek Database) và con số này còn tiếp tục tăng. Các chỉ thị CAPS được phát triển dựa trên công nghệ NGS SNP, dẫn đến việc hỗ trợ chọn lọc bằng chỉ thị phân tử rất hiệu quả ở lúa [31], [68], [102].

Sau hơn 20 năm phát triển chỉ thị CAPS, những thành công và hạn chế của phương pháp này trong sinh học thực vật hiện đã rõ ràng. Tính di truyền đồng trội của CAPS là một trong những đặc tính quan trọng nhất của loại chỉ thị này dẫn đến cả cá thể đồng hợp tử và dị hợp tử có thể dễ dàng được nhận ra trong quá trình đánh giá kiểu gen. Một ưu điểm quan trọng khác của CAPS là thiết bị đơn giản và tương đối rẻ. Như đã đề cập ở trên, thiết bị rất thông thường và phổ biến được sử dụng cho CAPS bao gồm PCR, cắt với enzyme giới hạn và điện di. Trong khi đó, hầu hết các kỹ thuật phân tử hiện đại khác đều dựa trên các chuỗi hoàn chỉnh hoặc một phần của bộ gen thực vật, yêu cầu thiết bị tự động cao và rất tốn kém [24], [75], [83], [111]. Ngoài ra, các kết quả đơn giản của CAPS gồm chỉ một hoặc một vài đoạn DNA sau cắt bằng enzyme giới hạn là một lợi thế rất tích cực của phương pháp này.

1.8. Đa hình nucleotit đơn (Single Nucleotide Polymorphism - SNP)

Chỉ thị DNA mới được gọi là đa hình nucleotit đơn trong hệ gen (SNP) gần đây đã và đang được ứng dụng rộng rãi trong các nghiên cứu về hệ gen. SNP thể hiện tính đa hình xảy ra bởi các đột biến điểm phát sinh trong các alen khác nhau bao gồm các bazơ thay thế tại vị trí nucleotit cụ thể. Sự khác biệt giữa các cá thể chính là những biểu hiện khác nhau về trình tự DNA. Gần đây, các nhà khoa học đã phát hiện được những sự khác biệt gây ra chỉ do sự

thay thế một nucleotit đơn, được gọi là sự đa hình nucleotit đơn (Single Nucleotide Polymorphisms - SNPs). Nghiên cứu gần đây trên tập đoàn giống lúa bản địa Việt Nam đã được phân tích kiểu gen nhờ gần 30.000 chỉ thị SNP phù hợp với phương pháp phân tích GWAS [88].

Nguyên lý phát hiện ra chỉ thị SNP bao gồm các bước chính sau: Sử dụng một cặp mồi được thiết kế dựa vào chỉ thị EST để nhân đoạn DNA đặc trưng cho một vùng giống nhau ở genom của mọi cá thể. Tiến hành xác định trình tự của sản phẩm nhân bản từ đó phát hiện ra sự đa hình do sự sai khác chỉ một nucleotit trong trình tự. SNP và trình tự liền kề có thể tìm được qua xây dựng thư viện và giải trình tự hoặc thông qua sàng lọc cơ sở dữ liệu trình tự có sẵn [54].

Tuy nhiên, trên thực tế chỉ một SNP đơn không thể nói lên nhiều. Để xác định được sự liên hệ giữa các SNP cần phải xem xét tổ hợp nhiều SNP trên một đoạn DNA dài hơn, mỗi tổ hợp SNP được gọi là một haplotype, trong một quần thể lớn, chỉ thấy một số tổ hợp SNP phổ biến tức là chỉ có một số kiểu haplotype nhất định tồn tại. Trong thế giới di truyền, mọi thông tin di truyền đều ở trạng thái cặp. Chúng ta sẽ có một haplotype từ mẹ, và một haplotype từ bố. Điều ấy có nghĩa là chúng ta có hai haplotype, hay một cặp haplotype. Hai haplotype trong cặp có thể khác biệt hoặc có thể giống nhau. Khi xem xét đến phản ứng của một cá thể đối với môi trường cần phải xét đến hồ sơ SNP đặc hiệu (SNP profile - cặp haplotype) của cá thể đó [4].

Mặc dù những tiến bộ công nghệ SNP đánh giá kiểu gen đã đạt được một số kết quả khả quan, tuy nhiên vẫn còn nhiều thách thức và đòi hỏi cần được có các thiết bị nghiên cứu chuyên dụng. Phương pháp truyền thống phổ biến sẵn có cho SNP đánh giá kiểu gen bao gồm: Giải trình tự trực tiếp, giải trình tự bazơ đơn, alen oligonucleotit cụ thể, biến tính trên gel điện di, SSCP và phản ứng chuỗi thắt (LCR). Ở thực vật, SNP nhanh chóng thay thế chỉ thị lặp

lại trình tự đơn giản. SNP là dạng phổ biến nhất của biến dị di truyền trong các hệ gen có nhân điển hình. Ứng dụng của SNP trong chọn giống cây trồng gồm: (1) Chọn lọc hệ gen và chọn lọc nhờ chỉ thị; (2) Lập bản đồ QTL và bản đồ tương quan, vị trí nhân bản [128]; (3) Lập bản đồ gen dựa trên phân tích các trình tự di truyền và phân tích phá hệ; (4) Kiểm tra độ thuần của giống (đặc biệt giống lai); (5) Xác định giống và (6) Giám sát alen kết hợp trong các môi trường cụ thể.

1.9. Tìm kiếm SNPs trong vùng genom mục tiêu bằng công nghệ chụp gen kết hợp với giải trình tự thế hệ mới.

Trước đây việc giải trình tự các giống sử dụng phương pháp Sanger rất tốn kém và mất thời gian, nhất là đối với các dự án nghiên cứu quy mô lớn, nhiều gen song song. Các phương pháp giải trình tự thế hệ mới (Next generation sequencing - NGS) đã giảm đáng kể chi phí và tăng sản lượng giải trình tự từ vài trăm nucleotit lên khoảng 600 tỷ bps mỗi lần chạy (ví dụ máy Illumina's HiSeq200 V3) [32], [124]. NGS được thực hiện bằng cách giải trình tự hàng loạt các đoạn DNA nhỏ theo kiểu song song. Do đó bước đầu tiên trong NGS là phân đoạn DNA bộ gen thành các mảnh nhỏ, thường là từ 300 - 500 bps bằng các công cụ cơ học hoặc hóa học [28]. Tiếp đó, các đoạn adapter được mã hóa khác nhau sẽ được thêm vào hai đầu của các phân đoạn DNA, cho phép gắn môi và giải trình tự phân đoạn DNA. Một đặc điểm chung của hầu hết các phương pháp NGS là các phân đoạn DNA được nhân dòng vô tính và sau đó được gắn lên các vi mạch (được gọi là Flow Cells, FlowChips hoặc PicoTiter), được sắp xếp theo kiểu song song.

Mặc dù các phương pháp giải trình tự thế hệ mới đã làm giảm đáng kể chi phí để giải trình tự toàn bộ bộ gen. Tuy nhiên đây chỉ là một phần nhỏ trong tổng chi phí dự án giải trình tự toàn bộ bộ gen trong đó bao gồm việc phân tích và quản lý dữ liệu giải trình tự. Ví dụ về giải trình tự toàn bộ bộ gen

người sẽ cần phải phân tích khoảng 3.2×10^9 nucleotit (bps). Trong khi đó số exon (toàn bộ vùng mã hóa protein) chỉ chiếm khoảng 1% bộ genôm, bao phủ 30 - 60 triệu bp [32]. Sử dụng công nghệ chụp gen (Gen capture) kết hợp với giải trình tự thế hệ mới giúp phân lập vùng cần giải trình tự trước, sau đó giải trình tự, là một giải pháp hữu ích trong trường hợp chỉ cần giải trình tự một vùng trong bộ genôm và do đó giúp làm giảm đáng kể lượng dữ liệu tạo ra cũng như công việc xử lý tin sinh, phân tích, quản lý dữ liệu.

Kết hợp các hệ thống chụp gen mục tiêu (Targeted Gene Capture) với giải trình tự thế hệ mới là phương pháp hiệu quả trong việc khám phá các vùng di truyền mục tiêu ở độ phân giải cao, cho phép khám phá nhanh chóng hàng ngàn đa hình di truyền và đa hình các nucleotit đơn (SNPs) từ đó giúp các nhà khoa học xây dựng chỉ thị phân tử phục vụ cho các mục tiêu nghiên cứu [46]. Phương pháp này có ưu điểm là 1) tiết kiệm chi phí đáng kể, 2) độ chính xác của trình tự cao hơn vì độ bao phủ có thể đạt được sâu hơn, 3) thời gian thực hiện ngắn hơn và 4) dữ liệu thu được phù hợp với các phân tích tin sinh xác định chức năng gen. Việc chụp gen mục tiêu kết hợp với NGS đã cho phép kiểm tra số lượng mẫu lớn hơn nhiều so với việc giải trình tự toàn bộ bộ gen [69].

Công nghệ chụp gen (Gen capture) gồm 4 bước: 1) Thiết kế mẫu dò bao phủ vùng cần giải trình tự sau đó mẫu dò được gắn với phân tử Biotin; 2) Lai các phân đoạn DNA với mẫu dò nhờ vào sự bắt cặp tương đồng; 3) Phân lập các phân đoạn DNA được lai với mẫu dò bằng các viên bi phủ streptavidine sẽ liên kết với Biotin trong mẫu dò; 4) Tách các phân đoạn DNA ra khỏi mẫu dò bằng cách tăng độ pH của dung dịch lai, làm bẻ gãy cầu nối giữa Biotin và Streptavidine, sẵn sàng để giải trình tự các phân đoạn DNA phân lập được trên máy giải trình tự thế hệ mới.

1.10. Nghiên cứu liên kết trên toàn hệ gen (GWAS), tiềm năng ứng dụng trong nghiên cứu về các tính trạng nông học phức tạp ở lúa

1.10.1. Phương pháp nghiên cứu liên kết trên toàn hệ gen GWAS

Mặc dù rất nhiều QTL và gen liên quan đến năng suất lúa đã được xác định bằng cách xây dựng bản đồ QTL, cung cấp thông tin chi tiết có giá trị về cơ sở di truyền của năng suất, nhưng những kết quả này chỉ giải thích một phần của đa dạng tự nhiên. Phương pháp nghiên cứu liên kết toàn bộ hệ gen (GWAS) ra đời đã nhanh chóng trở thành công cụ mạnh mẽ để xác định đa dạng alen trong tự nhiên vì nó so sánh đồng thời hiệu ứng của nhiều alen bằng cách sàng lọc một số lượng rất lớn các giống [23].

Trong số các ứng dụng công nghệ sinh học đã được áp dụng nghiên cứu trước đây ở lúa (nuôi cấy bao phấn, ưu thế lai, gây đột biến, cây trồng chuyển gen, lai chéo...), GWAS cùng với sự trợ giúp của các thể hệ giải trình tự mới ra đời cung cấp bộ chỉ thị phân tử bao phủ toàn bộ hệ gen với độ phân giải cao hoàn toàn chiếm ưu thế trong việc nghiên cứu các tính trạng phức tạp. GWAS cung cấp cái nhìn tổng quan về các tính trạng số lượng trong mối tương quan với kiểu gen, và các gen tiềm năng cho các nghiên cứu tiếp theo. Về nguyên tắc, GWAS đánh giá mối tương quan giữa mỗi chỉ thị di truyền với tính trạng nghiên cứu trong một quần thể cùng loài, cung cấp cơ sở di truyền của các tính trạng trên, cho phép lựa chọn các cặp bố mẹ tốt nhất để phân tích QTL cũng như đánh dấu các gen tiềm năng tác động đến tính trạng. Nếu các phương pháp nghiên cứu truyền thống thường chỉ có thể tập trung vào một vài tính trạng để nghiên cứu thì GWAS có thể áp dụng nghiên cứu trên cả tập đoàn cùng lúc nhiều tính trạng [51].

Lúa là một trong những cây trồng lý tưởng để tiến hành các nghiên cứu GWAS. Nghiên cứu GWAS ở lúa có nhiều thuận lợi vì bản chất của hầu hết các giống lúa là đồng hợp tử, do đó một khi quần thể nghiên cứu được đánh

giá đa dạng di truyền, dữ liệu di truyền có thể được tái sử dụng nhiều lần trên các kiểu hình và môi trường khác nhau. Với cơ chế tự thụ phấn nghiêm ngặt, độ phân rã của sự mất cân bằng di truyền liên kết (LD decay) giữa các chỉ thị di truyền qua các thế hệ thấp trong khi khoảng cách để sự mất cân bằng di truyền liên kết phân rã trong hệ gen lại cao hơn so với các cây trồng giao phấn [41], các dòng thuần có thể được duy trì qua nhiều thế hệ trong khi giới hạn độ phân giải để lập bản đồ gen liên kết lại nhỏ hơn so với các cây trồng giao phấn khác. Trong các nghiên cứu GWAS, với mật độ các SNP bao phủ toàn bộ hệ gen, khoảng tin cậy của các vùng QTL chỉ còn khoảng 50 - 100 kb, nhỏ hơn nhiều so với trong phương pháp lập bản đồ QTL trước đây, từ đó giúp cho việc tìm kiếm và xác định các gen tiềm năng dễ dàng hơn [51].

1.10.2. Ứng dụng nghiên cứu liên kết trên toàn hệ gen

GWAS được sử dụng lần đầu tiên hơn 10 năm trước trong di truyền ở người, đến nay đã có hơn 1.500 công bố về GWAS. GWAS được sử dụng khá phổ biến trong nghiên cứu các sinh vật mô hình, cũng như với cây trồng [71], [100] và đặc biệt là lúa gạo với hàng loạt các công bố trong khoảng 5 năm trở lại đây [18], [22], [43].

Hiện nay, đã có hơn 400 QTL liên quan đến tính trạng hình thái hạt đã được phát hiện bằng nhiều phương pháp lập bản đồ khác nhau [131] nhưng chỉ có gần 20 gen liên quan đến kích thước và khối lượng hạt được phân lập và nhân dòng. Trên tập đoàn giống lúa *Indica* Trung Quốc, Huang và cộng sự (2012) đã tìm thấy 32 lô-cut mới có liên quan đến thời gian ra hoa và 10 lô-cut liên quan đến kích thước hạt [50]. Trong khi đó trên tập đoàn giống lúa *Japonica* Trung Quốc, Si và cộng sự (2016) phát hiện một QTL lớn trên nhiễm sắc thể số 7 có liên kết với kích thước hạt và cấu trúc bông [97]. Năm 2015, với tập đoàn 369 giống lúa lai ưu tú (IRRI), Begum và cộng sự (2015) công bố 52 QTL mới liên quan đến 11 tính trạng nông học quan trọng gồm

nhiều QTL lớn có liên kết với các tính trạng thời gian ra hoa, chiều cao cây, năng suất, chiều dài, chiều rộng hạt [43]. Gần đây, Ma và cộng sự tiến hành một nghiên cứu GWAS liên quan đến kích thước hạt, trên 270 giống lúa, sử dụng 996722 SNP bao phủ trên toàn hệ gen, xác định được tương ứng 5 và 4 lô-cut có liên kết chặt với chiều dài và chiều rộng hạt, từ đây các nhà khoa học đã phát hiện gen *OsSNB* tham gia điều hòa âm tính đối với tính trạng kích thước hạt [74]. Đặc biệt, bằng việc sử dụng hơn 700000 SNP trên một tập đoàn 242 giống lúa nhiệt đới, Crowell và cộng sự đã phát hiện hai vùng QTL trên nhiễm sắc thể số 3 và 8 liên kết chặt với số lượng nhánh cấp một; một vùng trình tự trên nhiễm sắc thể số 9 cùng tác động đến chiều dài lóng và chiều dài nhánh cấp một; xác định được một vùng QTL trên nhiễm sắc thể 11 liên kết với chiều dài trục và chiều dài bông, đồng thời tìm kiếm thêm được nhiều vùng QTL khác có ảnh hưởng đến các tính trạng khác của cấu trúc bông [34]. Một nghiên cứu khác sử dụng hơn 300 giống lúa trên khắp thế giới, kết hợp với bộ chỉ thị phân tử hơn 44000 SNP, xác định được bảy QTL liên kết với số lượng nhánh cấp một, năm QTL liên kết với chiều dài bông, mười QTL liên kết với chiều dài hạt, trong số đó có một vùng QTL liên kết với số lượng nhánh cấp một có sự thống nhất giữa hai môi trường canh tác khác nhau [118].

Bên cạnh các tính trạng năng suất, tăng cường tính kháng ở lúa cũng là một thách thức lớn đối với các nhà chọn giống. Một nhóm các nhà khoa học Thái Lan thực hiện nghiên cứu GWAS trên quần thể 104 giống bản địa Thái Lan, với hơn 110000 SNP, xác định được bốn vùng QTL lớn có liên kết với tính trạng kháng mặn nằm trên các nhiễm sắc thể số 8, 12, 1 và 2 [64]. Gần đây, một nghiên cứu GWAS được tiến hành trên gần 200 giống bản địa Việt Nam, xác định được 17 QTL khác nhau liên kết với các đặc tính chịu hạn của lúa như hàm lượng nước tương đối, điểm nhạy cảm hạn, khả năng phục hồi sau stress... trong đó có 9 QTL biểu hiện sự liên kết với hai hay nhiều tính trạng [48].

Công bố gần đây nhất của Bai và cộng sự (2021) [16] cũng đã thực hiện nghiên cứu GWAS đối với ba tính trạng liên quan đến cấu trúc bông lúa (chiều dài bông - PL, số gié cấp một/bông - PBN, số gié cấp hai/bông - SBN) trên 340 giống lúa *Japonica* và *Indica* nằm trong dự án giải trình tự 3000 genom lúa có nguồn gốc từ 48 quốc gia thuộc Châu Á, Châu Âu, Châu Phi, Australia, Bắc và Nam Phi. Kết quả phân tích với 1087 SNP đã xác định được 129 QTL riêng lẻ (trong đó 115 QTL quy định tính trạng PL, 12 QTL quy định PBN và 2 QTL quy định SBN). Trong đó có 11 QTL thuộc nhóm *Japonica* (7 QTL quy định PL, 1 QTL quy định PBN và 3 QTL quy định SBN); 21 QTL thuộc nhóm *Indica* (6 QTL quy định PL, 6 QTL quy định PBN và 9 QTL quy định SBN). Chỉ 5 QTL (*qPL3-14/qPBN3-1*, *qPL5-3/qPBN5-1*, *qPL5-16/qPBN5-2*, *qPL6-1/qPBN6-1*, *qPL11-3/qSBN11-1*) quy định nhiều hơn một tính trạng và không có QTL nào cùng quy định với hai tính trạng PBN và SBN,

Tại Việt Nam, trong vài năm qua, một số nghiên cứu GWAS đã được tiến hành phân tích cho một loạt đặc điểm nông học ở lúa [48]. Trong dự án hợp tác Việt - Pháp, Phùng Phương Nhung và cộng sự đã xây dựng một quần thể các giống lúa bản địa đại diện cho các hệ sinh thái canh tác đa dạng ở Việt Nam gồm 182 giống lúa (115 *Indica* và 64 *Japonica*) được phân tích với gần 30000 SNP bằng kỹ thuật GBS (Genotyping By Sequencing) tạo tiền đề cho các nghiên cứu GWAS đối với các tính trạng cấu trúc rễ, cấu trúc bông.... Năm 2014, Phùng và cộng sự đã tiến hành phân tích GWAS trên các tính trạng liên quan đến cấu trúc bộ rễ (chiều dài rễ, trọng lượng rễ, độ dày rễ, số lượng rễ bên...), kết quả đã tìm ra hai QTL quan trọng và một số gen tiềm năng trên NST số 2 và số 11 ảnh hưởng đến tính trạng đường kính rễ và số lượng rễ bên [88], [89].

Tạ Kim Nhung và cộng sự (2018) đã tiến hành nghiên cứu GWAS các tính trạng liên quan đến cấu trúc bông của tập đoàn 159 giống lúa bản địa Việt Nam và đã xác định được 29 QTL ổn định có liên kết với bảy tính trạng liên quan cấu trúc bông, trong đó 6 QTLs quy định số hạt/bông, 9 QTLs quy định số gié/bông, 17 QTLs quy định các tính trạng chiều dài bông. Quan trọng hơn nghiên cứu này tìm ra sự có mặt của một vài vùng QTL mới liên kết với tính trạng số hạt/bông, số gié cấp hai và số gié cấp ba ở bông lúa. Đặc biệt nghiên cứu đã xác định được hai vùng QTL (QTL6 và QTL9) nằm trên nhiễm sắc thể số 2 cùng liên kết với hai tính trạng số hạt/bông và số gié cấp hai/bông. Tuy nhiên QTL6 chỉ chứa 1 SNP, QTL9 chứa 9 SNP bao phủ một vùng tương đối lớn trên nhiễm sắc thể số 2. Kết quả phân tích LD cũng chỉ ra đây là một vùng chứa nhiều SNP liên kết chặt chẽ với nhau. Hai QTLs khác là QTL14 và QTL19 cùng quy định cho tính trạng RL, PbinL và PBN, PBL. Không có QTLs nào cùng quy định cho PBL và SBL cũng như PBN và SBN [102]. Những kết quả này cung cấp nguồn dữ liệu giá trị cho các nghiên cứu di truyền ở lúa cũng như các chỉ thị phân tử có thể dùng trong chọn tạo lúa [51], [127].

1.11. Haplotype và phương pháp phân tích haplotype

Haplotype là sự kết hợp của các alen cho các dạng đa hình khác nhau (như SNP, chèn/xóa và các dấu hiệu hoặc biến thể khác) có trên cùng một nhiễm sắc thể, được di truyền cùng nhau, với sự mất cân bằng di truyền liên kết mạnh mẽ (LD) giữa chúng [99].

Trong quá trình tiến hóa của các loài cây trồng quan trọng như lúa, ngô, lúa mì... việc chọn lọc các gen/ alen quy định kiểu hình mong muốn cho tính trạng quan tâm là yếu tố chính dẫn đến việc hình thành các dấu hiệu của chọn lọc [91]. Các dấu hiệu của chọn lọc (còn được gọi là khối haplotype bảo tồn và quét chọn lọc) có nhiều gen, được quy định cùng nhau bởi nhiều gen điều hòa. Mối tương quan giữa các tính trạng khác nhau được phản ánh từ các dấu

hiệu chọn lọc là do mối liên kết giữa các gen hoặc do hiệu ứng đa hướng của các gen giống nhau. Do đó, các nhà chọn tạo giống cây trồng thường nhắm mục tiêu vào các vùng gen này để làm sáng tỏ ảnh hưởng của chúng đối với các tính trạng quan tâm. Bên cạnh đó, việc tích hợp bộ gen để xác định các tái tổ hợp được tạo ra bằng cách lai giữa các cặp bố mẹ tương phản sẽ hỗ trợ rất nhiều trong việc giải quyết sự phức tạp của các tính trạng số lượng [55].

Do sự sẵn có của dữ liệu giải trình tự từ số lượng lớn các cá thể cho một loài cây trồng nhất định, nên việc xác định haplotype đã trở nên dễ dàng hơn. Bằng cách sử dụng toàn bộ dữ liệu giải trình tự bộ gen, Bevan và các cộng sự (2017) đã xác định khái niệm về tập hợp haplotype. Cùng với dữ liệu kiểu hình của dòng mầm/dòng giống, có thể đánh giá và xác nhận các ảnh hưởng kiểu hình của các kiểu đơn bội 'thành phần' [25]. Dựa trên tiền đề này, và bằng cách sử dụng bộ dữ liệu giải mã lại toàn bộ bộ gen quy mô lớn kết hợp với phân tích kiểu hình haplotype, Abbai và các cộng sự (2019) đã xác định được các haplotype hữu ích cho việc nhân giống lúa trong tương lai [9] và Sinha và các cộng sự đã làm theo cách tiếp cận tương tự ở cây đậu triều (*Pigeon pea*) [98]. Dữ liệu SNP mật độ cao được tạo ra từ nhiều kiểu gen thông qua các phương pháp tiếp cận dựa trên NGS hoặc dựa trên mảng đã được sử dụng để phát triển các dạng đơn bội ở nhiều loài thực vật. Các đơn bội này cũng đã được sử dụng cho các ứng dụng khác nhau trong nghiên cứu và nhân giống ở các loài cây trồng khác nhau.

CHƯƠNG II

VẬT LIỆU, NỘI DUNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu và các thiết bị trong nghiên cứu

2.1.1. Vật liệu nghiên cứu

- Bộ dữ liệu phân tích hình thái cấu trúc bông và các tính trạng liên quan đến năng suất lúa của tập đoàn 155 giống lúa bản địa Việt Nam được kế thừa từ kết quả nghiên cứu GWAS của nhóm Khổng Ngân Giang, Tạ Kim Nhung và các cộng sự năm 2014 - 2015 (công bố năm 2018) phục vụ phân tích, chọn lọc các giống lúa bố mẹ để lai tạo quần thể F_1 [102].

- Bốn giống lúa bản địa Việt Nam (cung cấp bởi Trung tâm Tài nguyên thực vật) làm vật liệu lai tạo, được chọn lọc từ tập đoàn 155 giống lúa trên, thuộc hai haplotype của QTL9, có kiểu hình tương phản về cấu trúc bông. Haplotype 1 gồm hai giống Sớm Giai Hưng Yên (G6), Ôn (G19), có cấu trúc bông nhỏ, chứa 9 SNPs (gagagcgaa) sử dụng làm giống mẹ. Haplotype 2 gồm hai giống Khẩu Nam Rinh (G189), Blé Blâu Cho (G205) có cấu trúc bông to, chứa 9 SNPs (atataaatt) sử dụng làm giống bố. Thông tin chi tiết về các giống được trình bày trong bảng 2.1.

- 12 cặp mồi SSR (Microsatellite Marker) đã được công bố cho sự đa hình giữa các giống lúa, nằm trên các nhiễm sắc thể khác nhau được sử dụng để chọn lọc các cây F_1 (Bảng 2.2)

- Trình tự gen giống lúa tham chiếu Nipponbarre [58], [77].

2.1.2. Thiết bị sử dụng trong nghiên cứu

Các thiết bị máy móc sử dụng trong sinh học phân tử: Máy điện di, máy đo nồng độ DNA Nanodrop 8000, hệ thống điện di mao quản Qiaxcel, máy PCR, máy ly tâm, buồng soi gel UV (Uvltec, Cambridge) phần mềm Fire Reader, máy cắt DNA siêu âm tập trung, máy giải trình tự thế hệ mới Illumina

Miseq en pair end (2x250), máy điện di Fragment Analyser...và các loại hóa chất chuyên dụng sử dụng trong tách chiết DNA, chạy phản ứng PCR.

Bảng 2.1. Một số chỉ số về cấu trúc bông và năng suất của các giống lúa sử dụng làm bố mẹ để tạo các quần thể lai F₁

(Tại Trung tâm Tài nguyên Di truyền thực vật vụ mùa 2014 và Trạm Nghiên cứu Nông nghiệp Văn Giang vụ mùa năm 2015)

Tên giống	Địa điểm thu thập	Năm	Chiều dài bông (cm)	Số gié cấp 1/ bông	Chiều dài gié cấp 1 (cm)	Khoảng cách giữa các gié cấp 1 (cm)	Số gié cấp 2/ bông (gié)	Chiều dài gié cấp 2 (cm)	Khoảng cách giữa các gié cấp 2 (cm)	Số gié cấp 3/ bông (gié)	Số hạt/ bông (hạt)	Thời gian ra hoa (ngày)
	Không xác định											
G6	xác định	2014	18,21	11,22	11,88	1,81	23,67	2,93	1,12	0,06	130,22	96,5
G19	Hà Nội	2014	19,61	11,39	12,10	1,90	24,22	2,58	1,16	0,0	133,94	95,5
G189	Điện Biên	2014	20,41	11,83	14,29	1,91	48,67	3,30	0,72	0,44	250,50	94,5
G205	Sơn La	2014	22,09	11,78	14,58	2,05	44,44	3,15	0,99	0,0	220,33	99,0
	Không xác định											
G6	xác định	2015	25,04	13,61	16,38	2,03	34,83	4,01	1,40	0,06	187,33	101,0
G19	Hà Nội	2015	27,58	13,83	17,02	2,18	31,11	3,53	1,69	0,44	176,83	100,5
G189	Điện Biên	2015	29,85	12,72	20,59	2,57	45,22	4,76	1,19	0,89	228,22	79,0
G205	Sơn La	2015	28,97	13,33	18,30	2,35	47,39	3,97	1,38	0,50	244,22	108,0

Bảng 2.2. Danh sách 12 chỉ thị phân tử SSR sử dụng để chọn lọc cây F₁

Tên chỉ thị	Số hiệu	NST	Trình tự lặp	Kích cỡ (bp)	Trình tự mồi	Tác giả
RM151	L37528	1	(TA)23	205 - 317	F-ggctgctcatcagctgcatgcg R-tcggcagtgtagagtttgatctgc	Temnykh và cộng sự, 2000 Akagi và cộng sự, 1996
RM180	D63901	7	(ATT)10	107 - 204	F-ctacatcggcttaggttagcaacacg R-acttgctctactgtggtagggactg	Temnykh và cộng sự, 2000
RM204	AF344025	6	(CT)44	106 - 194	F-gtgactgactggatcataggg R-gctagccatgctctcgtacc	Chen và cộng sự, 1997
RM289	AF344115	5	G11(GA) 16	88 - 180	F-ttccatggcacacaagcc R-ctgtgcacgaactccaag	Temnykh và cộng sự, 2000
RM320	AF344145	7	(AT)11GT AT(GT)13	153 - 254	F-caacgtgatcgaggatagatc R-ggatttgcttaccacagctc	Temnykh và cộng sự, 2000
RM400	AQ051253	6	(ATA)63	195 - 321	F-acaccaggetacccaaactc R-cggagagatctgacatgtgg	Temnykh và cộng sự, 2001
RM410	AQ156440	9	(TA)13	173 - 269	F-gctcaacgtttcgttcctg R-gaagatgcgtaaagtgaacgg	Temnykh và cộng sự, 2001
RM491	AQ510175	12	(AT)14	263 - 400	F-acatgatgcgtagcgagttg R-ctctcccttccaattctc	Temnykh và cộng sự, 2001
RM532	AQ843286	3	(CA)9	166 - 180	F-tctataatgtagccccccc R-tttcaggggcttctaccaac	Temnykh và cộng sự, 2001
RM535	AQ857127	2	(AG)11	138 - 410	F-actacatacacggcccttgc R-ctacgtggacaccgtcacac	Temnykh và cộng sự, 2001
RM577	AP000816	1	(TA)9(CA) 8	191 - 270	F-gctttccctctaaccctct R-ggatgtaccgctgacatgaa	Temnykh và cộng sự, 2001 McCouch và cộng sự, 2008
RM592	AC016779	5	(ATT)20	210 - 400	F-tctttggtatgaggaacacc R-agagatccggtttgtgtaa	Temnykh và cộng sự, 2001

Ghi chú: NST: Nhiễm sắc thể

2.2. Địa điểm và thời gian nghiên cứu

- Địa điểm thực hiện:

+ Thí nghiệm phân tích, đánh giá kiểu gen QTL9 các quần thể F_1 , F_2 , F_3 thực hiện tại Viện Di truyền Nông nghiệp (phường Cổ Nhuế 1, quận Bắc Từ Liêm, thành phố Hà Nội);

+ Thí nghiệm chụp gen kết hợp giải trình tự Illumina và phân tích dữ liệu SNP, phát triển chỉ thị phân tử CAPS thực hiện tại Viện Nghiên cứu và phát triển IRD - Pháp;

+ Thí nghiệm đánh giá kiểu hình thực hiện ngoài đồng ruộng tại xã Đông La, huyện Hoài Đức, thành phố Hà Nội.

- Thời gian thực hiện luận án: Từ năm 2017 đến năm 2020.

2.3. Nội dung nghiên cứu

2.3.1. Nội dung 1: Chọn lọc các cặp lai bố mẹ và lai tạo quần thể F_1

2.3.1.1. Phân tích haplotype vùng QTL9 của tập đoàn lúa sử dụng trong nghiên cứu GWAS, chọn lọc các giống lúa bố mẹ làm vật liệu lai tạo

2.3.1.2. Tạo quần thể lai F_1 và chọn lọc cây F_1 bằng chỉ thị phân tử SSR

2.3.2. Nội dung 2: Phát triển chỉ thị phân tử CAPS để chọn lọc các cây F_2 mang QTL9 đồng hợp thuộc hai haplotype bố hoặc mẹ

2.3.2.1. Xác định các SNPs trong vùng QTL9 của hai giống bố mẹ để phát triển chỉ thị phân tử CAPS

2.3.2.2. Phát triển chỉ thị phân tử CAPS

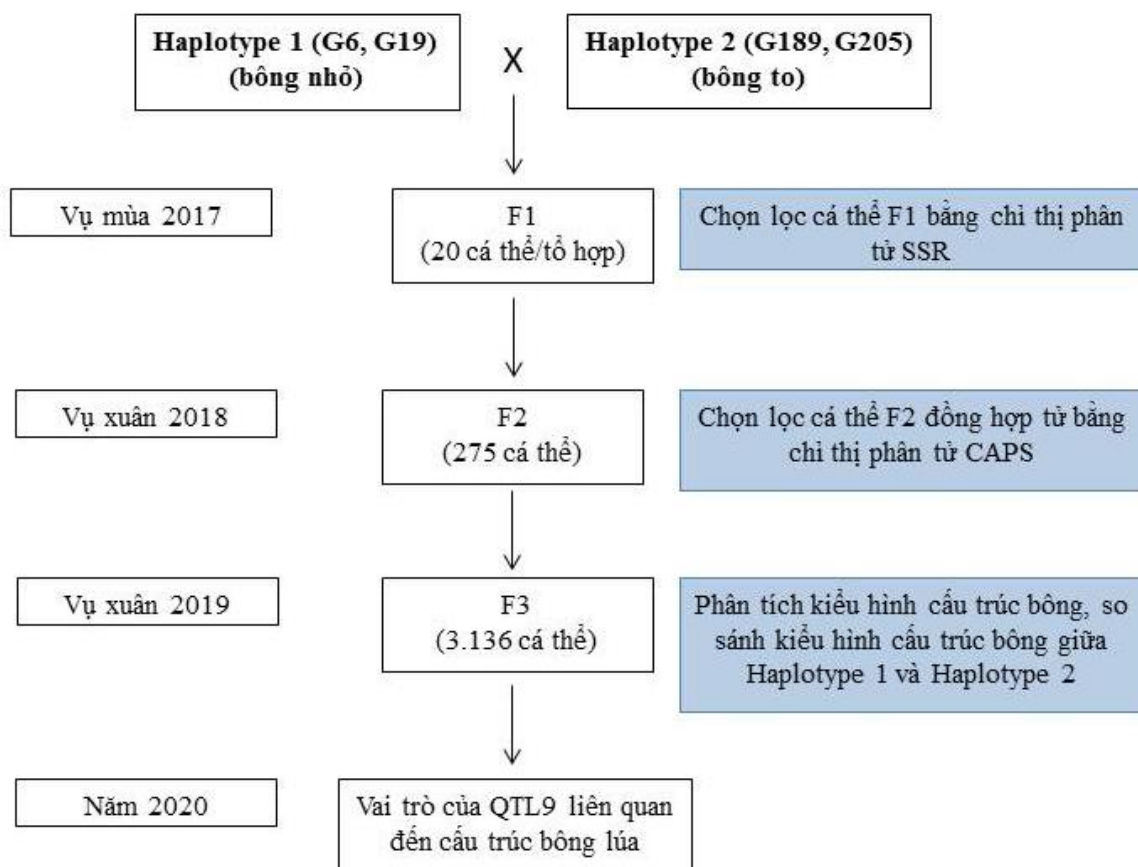
2.3.3. Nội dung 3: Chọn lọc các cây F_2 mang QTL9 đồng hợp thuộc hai haplotype bố hoặc mẹ bằng chỉ thị phân tử CAPS

2.3.3.1. Xác định chỉ thị phân tử CAPS cho đa hình giữa hai giống lúa bố mẹ

2.3.3.2. Chọn lọc các cây F_2 đồng hợp tử mang QTL9 của bố hoặc mẹ bằng chỉ thị phân tử CAPS

2.3.4. Nội dung 4: Phân tích kiểu hình cấu trúc bông của hai quần thể F_3 thuộc hai haplotype bố hoặc mẹ

Sơ đồ tóm tắt các nội dung nghiên cứu thể hiện trong hình 2.1



Hình 2.1. Sơ đồ lai và chọn lọc MAS để đánh giá vai trò của QTL9 liên quan đến cấu trúc bông lúa

2.4. Phương pháp nghiên cứu

2.4.1. Phân tích haplotype vùng QTL9 của tập đoàn lúa sử dụng trong nghiên cứu GWAS và xác định các giống lúa bố mẹ để tạo quần thể lai F_1

Kế thừa kết quả phân tích hình thái cấu trúc bông và các tính trạng liên quan đến năng suất lúa của tập đoàn 155 giống lúa bản địa Việt Nam nằm trong nghiên cứu GWAS của nhóm Khổng Ngân Giang, Tạ Kim Nhung và các cộng sự năm 2014 - 2015 [102]. Tiến hành phân tích haplotype sử dụng các gói công cụ trong phần mềm R (R software) theo Morandat và các cộng

sự (2012) [80].

Số lượng haplotypes trên tập đoàn lúa bản địa Việt Nam được tính toán bằng cách sử dụng gói *haplotypes* và lệnh R ngắn theo Aktas (2020) [11], sử dụng các chức năng trong gói *pegas* theo Paradis (2010) [84] để tính khoảng cách giữa các haplotype và xây dựng mạng lưới haplotype bằng cách chia các màu khác nhau.

Từ đó chọn lọc các giống thuộc hai haplotype có các tiêu chí sau:

- Đa hình về 9 SNP (gagagcgaa/atataaatt) nằm trong vùng QTL9
- Có kiểu hình cấu trúc bông khác biệt về số gié cấp hai/bông và số hạt/bông
- Thuộc cùng nhóm (*Indica* hoặc *Japonica*)
- Thời gian sinh trưởng gần nhau.

2.4.2. Phương pháp tạo các quần thể lai

Quần thể F_1 được lai tạo bằng phương pháp lai truyền thống và tự thụ các cá thể F_1 , F_2 để tạo quần thể F_2 , F_3 .

Các giống lúa được chọn làm bố mẹ để lai tạo quần thể F_1 sẽ được gieo trồng tại nhà lưới Viện Di truyền Nông nghiệp, mỗi giống 5 cây, trồng thành 3 đợt, mỗi đợt cách nhau 1 tuần. Cây được trồng trong xô nhựa có dung tích 5 lít chứa 50% đất thịt + 50% giá thể chuyên dùng trồng lúa. Các giống có thời gian sinh trưởng dài ngày hơn được trồng sớm hơn 1 - 2 tuần so với giống có thời gian sinh trưởng ngắn hơn. Sau 2 tháng gieo trồng tiến hành xử lý quang chu kỳ để điều khiển ra hoa: 9 giờ chiếu sáng trong ngày (từ 8 giờ sáng đến 17 giờ chiều) và 15 giờ trong buồng tối, lặp lại hàng ngày cho đến khi cây lúa trở bông.

Mô tả cách lai lúa để tạo hạt F_1 :

Lúa có đòng lớn đến giai đoạn trổ bông. Khi bông lúa thoát ra khỏi sự bao bọc của lá và dần tách rời lá dinh dưỡng để hở cổ bông cùng với trên đỉnh bông có một số hoa đã bắt đầu bung phần là thời điểm thích hợp nhất để lai.

- Dùng tay tách nhẹ nhàng bông lúa và lá gần nhất, cắt bớt 2/5 lá này
- Dùng kéo cắt bớt các nhánh non từ cổ bông lên và các nhánh có hoa đã bung phần làm sao để chỉ giữ lại 3/5 số hoa trên bông là những hoa đã chín.
- Cắt chóp hoa cây mẹ dùng mũi kéo loại bỏ toàn bộ hạt phấn (5 hạt).
- Rũ bông của cây bố bên trên vị trí bông của cây mẹ đã loại hạt phấn để cho túi nhụy của cây bố rơi vào hoa đã loại nhụy của cây mẹ.
- Chụp bông lúa của cây mẹ bằng bao giấy.
- Sau 4 đến 5 ngày thì nhìn rõ hạt F_1 phát triển.
- Khi thấy rõ hạt F_1 bỏ chụp cho bông phát triển bình thường.

2.4.3. Chọn lọc cây F_1 bằng chỉ thị phân tử SSR

2.4.3.1. Thu thập mẫu và tách chiết DNA

Thu mẫu lá non 20 ngày tuổi (lấy lá thứ 2 từ trên xuống) của 20 cây F_1 của mỗi cặp lai. Tách chiết DNA theo phương pháp sử dụng CTAB của Doyle có cải tiến (1991) [40]. Đo nồng độ và kiểm tra độ tinh sạch của DNA bằng máy Nanodrop 8000. Kiểm tra chất lượng DNA bằng cách điện di trên gel Agarose 1% (100 V; $0,5 \times$ TBE, 30 phút), sử dụng thang chuẩn DNA Gene Ruler 1kb (#SM1163, Fermentas).

2.4.3.2. Phản ứng PCR với các cặp mồi SSR

Phản ứng PCR được thực hiện trong thể tích 20 μ l, gồm: 20 ng DNA; 0,1 mM dNTPs (#R0182, Fermentas); 0,1 μ M mồi SSR xuôi và ngược (bảng 2.2); 2 μ l 10X Dream Taq Buffer (#R0971, Fermentas); 0,2 μ l Dream Taq DNA polymerase 5u/ μ l (#EP0702, Fermentas) và nước Milli Q khử trùng.

Chu trình nhiệt: 95°C trong 5 phút, 30 chu kỳ; 95°C trong 30 giây, 65°C trong 30 giây (nhiệt độ gắn mồi), 72°C trong 30 giây và giữ 5 phút ở

72°C. Sản phẩm PCR được kiểm tra bằng điện di trên gel agarose 2,5% rồi được soi trong buồng UV (Uvltec, Cambridge), phần mềm Fire Reader. Các thí nghiệm điện di sử dụng thang chuẩn DNA Low Gene Ruler 100 bp (#SM1163, Fermentas).

2.4.4. Xác định các SNPs trong vùng QTL9 ở các giống bố mẹ bằng phương pháp chụp gen (Gene capture) kết hợp với giải trình tự thế hệ mới

Sử dụng công nghệ chụp gen kết hợp với giải trình tự Illumina, bao gồm các bước sau:

Bước 1: Xây dựng thư viện mẫu dò RNA biotin hóa bao phủ vùng QTL9;

Bước 2: Xây dựng thư viện mẫu DNA;

Bước 3: Lai mẫu dò và DNA vùng mục tiêu;

Bước 4: Giải trình tự vùng QTL9 bằng hệ thống giải trình tự thế hệ mới Illumina;

Bước 5: Phân tích và sàng lọc dữ liệu, xác định các SNPs nằm trong vị trí cắt của enzyme giới hạn để xây dựng chỉ thị phân tử CAPS.

2.4.4.1. Bước 1: Xây dựng thư viện mẫu dò RNA biotin hóa bao phủ vùng QTL9

Vùng QTL9 chứa 137 gen, mẫu dò được thiết kế bao phủ toàn bộ 137 gen, kể cả vùng promoter (1000 pb trước mã mở đầu) và vùng 3' prime (500 pb sau mã kết thúc). Mẫu dò được thiết kế bằng phần mềm Mycroarray và sử dụng trình tự của giống lúa tham chiếu Nipponbare. Mỗi đoạn mẫu dò có chiều dài 80 pb, trong đó 40 pb trùng với mẫu dò kế tiếp. Các trình tự mẫu dò sau khi thiết kế xong, được tổng hợp bởi công ty DNAid (Pháp) tạo thành một thư viện các sợi DNA oligonucleotit. Các sợi đơn DNA sau đó được phiên mã «*in vitro*» thành RNA và được biotin hóa.

2.4.4.2. Bước 2: Xây dựng thư viện mẫu DNA

Sử dụng 1 µg DNA tổng số của mỗi giống lúa được cắt thành những mảnh nhỏ DNA có độ dài trung bình khoảng 350 bp bằng máy cắt siêu âm tập trung. Độ dài các mảnh DNA được kiểm tra bằng máy điện di Fragment Analyser. DNA được tinh sạch bằng các hạt từ tính Agencourt AMPure XP SPRI (Beckman Coulter, Úc), tiếp theo là sửa chữa kết thúc và cắt đuôi A, sau đó được gắn với adaptator và index có trình tự đặc thù của Illumina p5 (CACTGC) và p7 (GCGCTA) (IDT × Gen Custom Blocking Oligos). Phản ứng PCR được thực hiện để làm giàu các đoạn DNA mục tiêu, sử dụng cặp mồi đặc hiệu của Illumina. Để kiểm tra chất lượng của quá trình làm giàu mục tiêu, phản ứng qPCR được thực hiện với cặp mồi đặc hiệu cho mỗi vùng mục tiêu. Cuối cùng tất cả các mẫu DNA mục tiêu được gộp chung lại với nhau để phục vụ việc giải trình tự.

2.4.4.3. Bước 3: Lai mẫu dò và DNA vùng mục tiêu

Quá trình lai mẫu dò và DNA vùng mục tiêu được thực hiện bằng cách sử dụng bộ kit homemade của hãng DNAid, theo phương pháp « MYBAITs » phiên bản 2 của MYcroarray (“MyBaits - Hyb Capture Kits”) [82] (hình 2.2) bao gồm 5 bước:

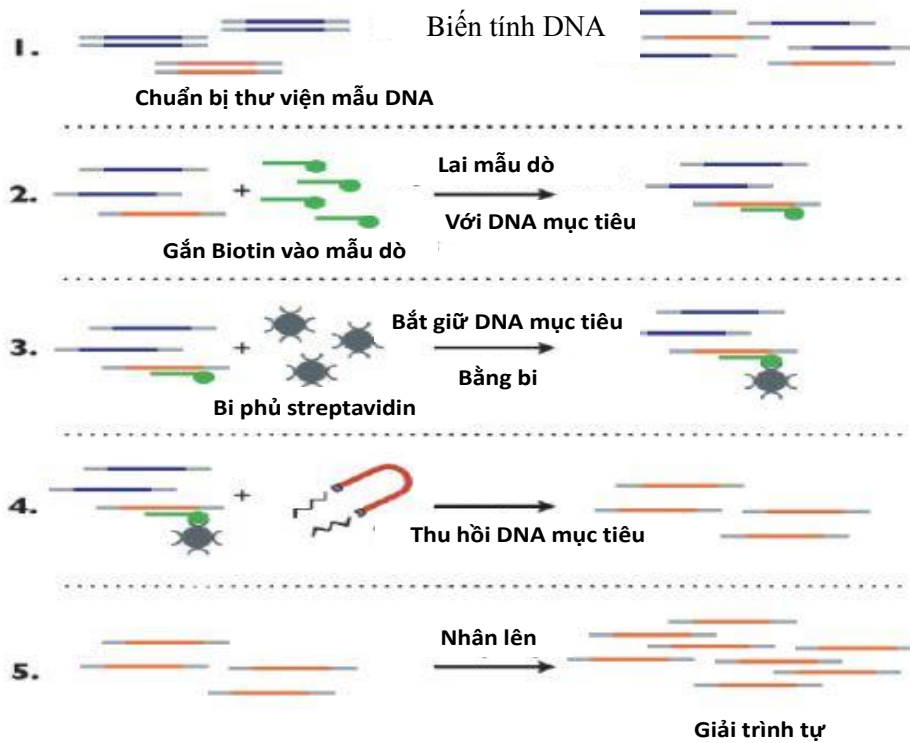
(1) Biến tính DNA mục tiêu: các mẫu DNA được biến tính ở 95°C trong 5 phút để giãn mạch, tạo thành các DNA mạch đơn.

(2) Lai mẫu dò với DNA mục tiêu: Các mẫu dò RNA biotin hóa được lai với DNA mục tiêu mạch đơn trong dung dịch lai, ở 65°C trong 36 giờ.

(3) Hỗn hợp lai mẫu và đầu dò được gắn vào các hạt từ tính streptavidin.

(4) Thu hồi các đoạn DNA mục tiêu: Các phức lai DNA - RNA sau đó được giữ lại sau khi loại bỏ các đoạn DNA không phải là mục tiêu, tách khỏi đầu dò và được làm giàu.

(5) Nhân lên: Các đoạn DNA được nhân lên bằng PCR và sau đó được giải trình tự.



Hình 2.2. Quá trình lai mẫu dò và DNA vùng mục tiêu (theo MyBaits V2 của MYcroarray)

2.4.4.4. Bước 4: Giải trình tự thư viện mẫu

Giải trình tự của thư viện mẫu được thực hiện trên một lần của máy giải trình tự thế hệ mới Illumina Miseq en pair end (2x250)

2.4.4.5. Bước 5: Phân tích và sàng lọc dữ liệu, xác định các SNPs nằm trong vị trí cắt của enzyme giới hạn để xây dựng chỉ thị phân tử CAPS.

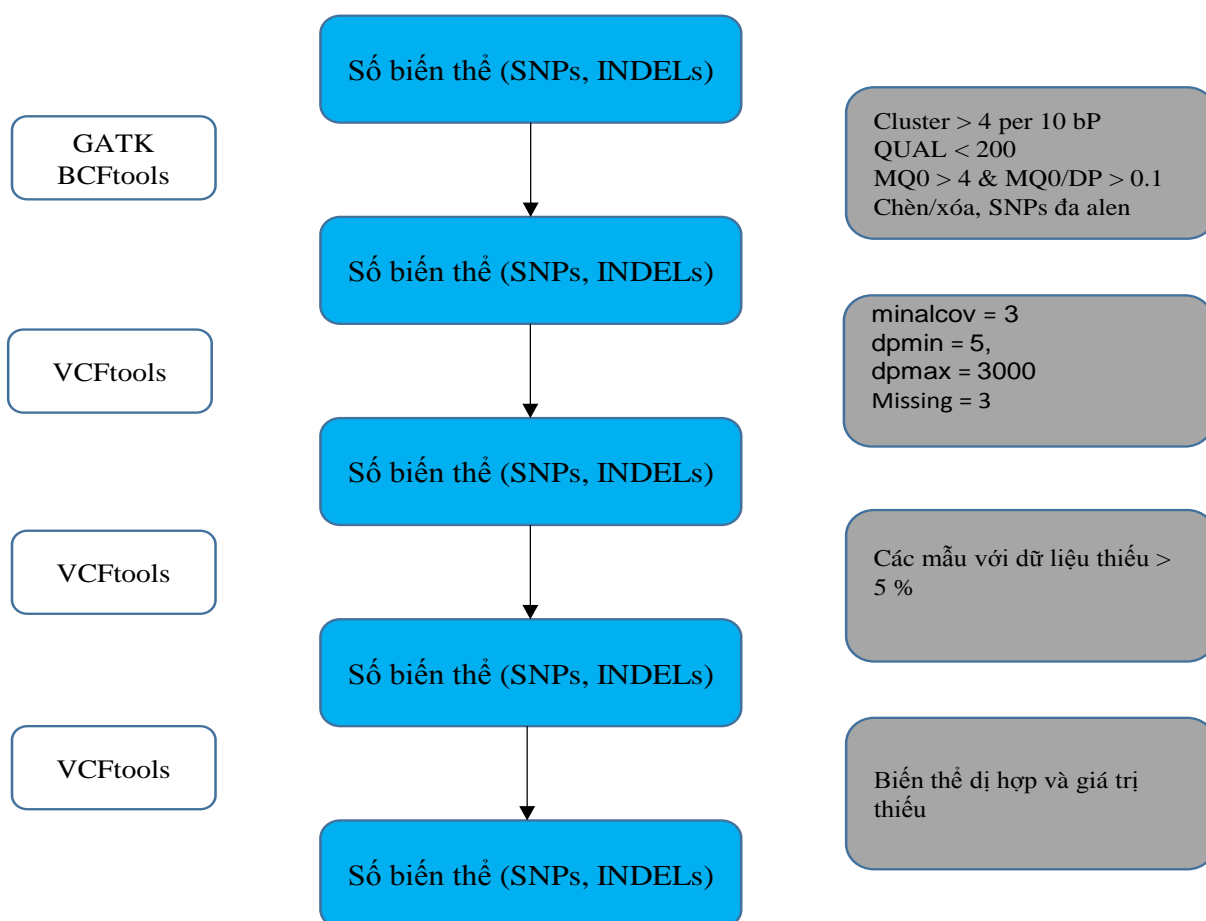
**Gọi biến thể*: Dữ liệu trình tự đọc thô được đánh giá, kiểm tra chất lượng và nhận diện các lỗi trong dữ liệu bằng công cụ FastQC. Sau đó dữ liệu sẽ được giống hàng với bộ gen tham chiếu và loại bỏ vị trí phân tử trùng lặp, sử dụng phần mềm BWA mem. Kết quả giống hàng dữ liệu sẽ được sàng lọc bằng phần mềm SAMtools để đảm bảo các đoạn trình tự đọc đã được giống

hàng chính xác với bộ gen tham chiếu và loại bỏ những sai sót. Các biến thể (SNPs, INDELS) sẽ được phát hiện bằng công cụ Haplotypecaller (hình 2.3)



Hình 2.3. Các bước gọi biến thể sử dụng TOGGLE

* *Sàng lọc biến thể*: Các dữ liệu đọc thô được ánh xạ với trình tự bộ gen tham chiếu theo Kawahara và cộng sự (2013) [58] và được xử lý để chọn lọc ra các biến thể bằng công cụ TOGGLE theo Monat và cộng sự (2015) [79] (hình 2.3). Sau đó được lọc để lựa chọn SNPs đồng hợp và loại bỏ SNPs dị hợp và trùng lặp. Quá trình lọc được thực hiện bằng cách sử dụng VCFtools 0.1.16 theo Danecek và cộng sự (2011) [35], SAMtools 1.9 theo Li và cộng sự (2009) [66] và GATK 4.0.0.0 theo DePristo và cộng sự (2011) [36], Van der Auwera và cộng sự (2013) [110]. Các SNPs chất lượng và biallelic được giữ lại cho các phân tích tiếp theo. Tiêu chí lọc chi tiết được mô tả trong hình 2.4.



Hình 2.4. Các bước sàng lọc biến thể và các tiêu chí sàng lọc tương ứng với từng bước

(Các công cụ sàng lọc trong hộp màu trắng, tiêu chí sàng lọc màu ghi)

Sàng lọc biến thể được thực hiện theo các tiêu chí sau:

- Số lần đọc tối thiểu để gắn 1 biến thể dị hợp vào một giống (-minalcov = 3).
- Độ sâu tối thiểu (5 lần đọc) và độ sâu tối đa (3000 lần đọc) được phép để gắn một kiểu gen cho một mẫu (dpmin 5, dpmax 3000)
- Số lượng mẫu tối đa chứa một kiểu gen bị thiếu (missing = 3)

* *Chú giải biến thể:* Việc chú giải biến thể được thực hiện bằng phần mềm SnpE. Các bước chú giải biến thể và chuyển hóa các file được thực hiện bằng công cụ Galaxy của platforme Southgreen, sử dụng SNIPlay3, một ứng dụng trực tuyến để khám phá và phân tích quy mô lớn về các biến thể gen theo Dereeper và cộng sự (2011) [37].

2.4.5. Phát triển chỉ thị phân tử CAPS

* *Nguyên tắc của chỉ thị CAPS*: là dựa vào sự đa hình của SNP trong vùng trình tự cắt của enzyme giới hạn giữa các cá thể khác nhau. Ở những cá thể xảy ra đột biến trong vị trí cắt của enzyme giới hạn (thường là SNPs) thì enzyme sẽ không nhận ra để cắt. Từ đó, mỗi được thiết kế để nhận đoạn DNA chứa vị trí cắt của enzyme giới hạn. Sản phẩm PCR sau đó được cắt bởi enzyme giới hạn sẽ cho các đoạn DNA có chiều dài khác nhau giữa giống có đột biến trong vị trí cắt của enzyme giới hạn (chỉ có 1 đoạn DNA dài vì enzyme không cắt sản phẩm PCR) và giống không có đột biến (sản phẩm PCR được cắt thành 2 mảnh DNA ngắn hơn). Đa hình của sản phẩm cắt bằng enzyme giới hạn dễ dàng quan sát được trên gel agarose cho phép phân biệt các giống khác nhau. Tính di truyền đồng trội của CAPS là một trong những đặc tính quan trọng nhất của loại chỉ thị này dẫn đến cả cá thể đồng hợp tử và dị hợp tử có thể dễ dàng phân biệt được.

* *Các bước phát triển chỉ thị phân tử CAPS như sau*:

- Thiết kế mỗi để nhận đoạn DNA chứa SNPs sao cho sản phẩm PCR khi bị cắt bởi enzyme giới hạn sẽ thành 2 mảnh có chiều dài khác nhau, phân biệt được trên gel agarose. Các mỗi thiết kế dựa trên trình tự của giống Nipponbarre trên trang MSU Rice Genome Annotation (Osa1) Release 7 (<http://rice.plantbiology.msu.edu>), sử dụng phần mềm Primer3 plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) [137].

- Thử nghiệm độ đặc hiệu của mỗi và sự đa hình của CAPS ở các giống khác nhau bằng phản ứng PCR và cắt enzyme giới hạn, sau đó điện di sản phẩm cắt trên gel agarose.

2.4.6. Chọn lọc các cây F_2 mang QTL9 đồng hợp thuộc hai haplotype bố hoặc mẹ bằng chỉ thị phân tử CAPS

2.4.6.1. *Tách chiết DNA*: DNA tổng số được tách chiết từ mẫu lá lúa 20 - 25 ngày tuổi theo phương pháp CTAB của Doyle (1991) có cải tiến [40] như trình bày ở trên.

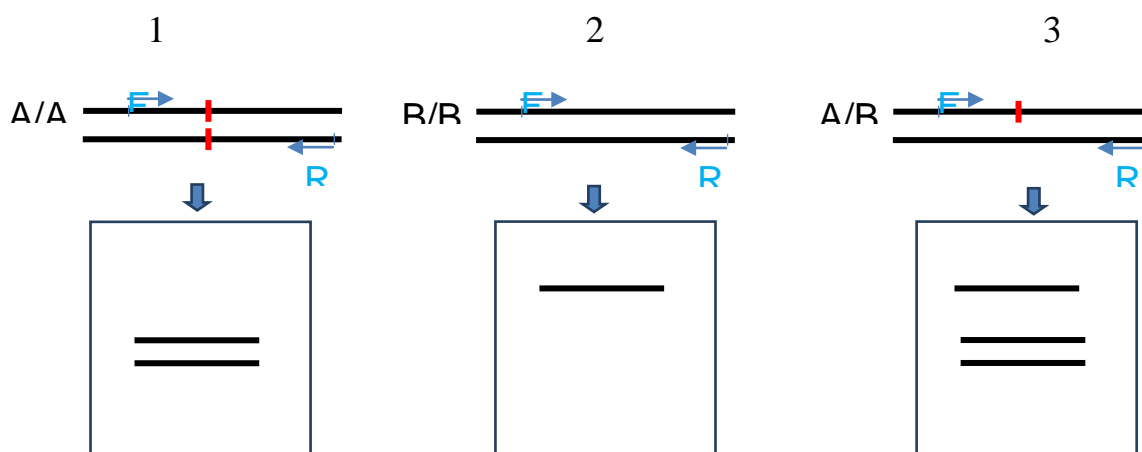
2.4.6.2. *Phản ứng PCR với cặp môi CAPS*

- Phản ứng PCR được thực hiện trong thể tích 10 μ l, gồm: 1,25 μ l DNA 20 ng; 0,2 μ l dNTPs 10X (#R0182, Fermentas); 0,5 μ l môi CAPS xuôi và ngược; 0,8 μ l $MgCl_2$; 2 μ l GoTaq Buffer 10X (#R0971, Fermentas); 0,05 μ l GoTaq DNA polymerase 5u/ μ l (#EP0702, Fermentas) và 5,2 μ l nước Milli Q khử trùng.

- Chu trình nhiệt: 95°C trong 2 phút, 35 chu kỳ, 95°C trong 30 giây, 57°C trong 30 giây (nhiệt độ gắn môi), 72°C trong 1 phút và giữ 7 phút ở 72°C. Sản phẩm PCR được kiểm tra bằng điện di trên gel agarose 2,5% rồi được soi trong buồng UV (Uvltec, Cambridge), phần mềm Fire Reader. Các thí nghiệm điện di sử dụng thang chuẩn DNA Low Gene Ruler 100 bp (#SM1163, Fermentas).

2.4.6.3. *Cắt sản phẩm PCR bằng enzyme giới hạn tương ứng*

Các phản ứng cắt sản phẩm PCR bằng enzyme giới hạn được thực hiện với các kit của hãng Thermo Fisher Scientific bao gồm enzyme và dung dịch đệm thích hợp theo hướng dẫn của nhà sản xuất, mỗi phản ứng được thực hiện trong thể tích 20 μ l bao gồm DNA, enzyme giới hạn và dung môi. Các phản ứng cắt trong đề tài này sử dụng các enzyme *EcoRI*, *BamHI*, *Sall*, *DraI* hoạt động tối ưu ở 37°C. Hỗn hợp phản ứng được ủ ở 37°C qua đêm, sau đó được kiểm tra bằng cách điện di trên gel agarose 2% (hình 2.5).



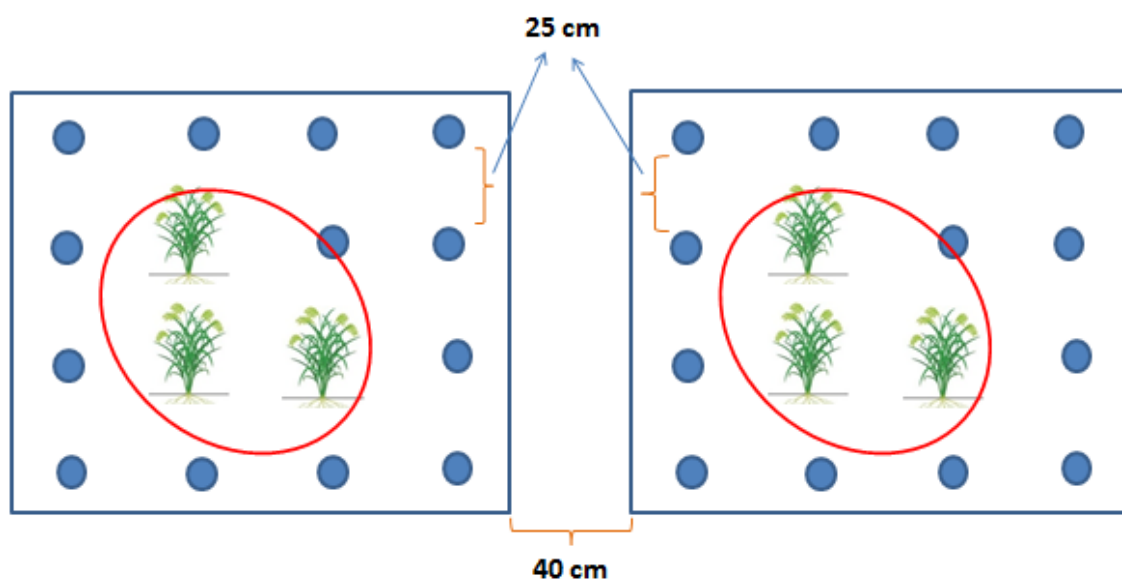
Hình 2.5. Mô hình sử dụng chỉ thị CAPS xác định cá thể đồng hợp tử

(1 - cá thể đồng hợp tử A/A không chứa SNP nằm trong trình tự cắt của enzyme giới hạn nên bị cắt, sản phẩm điện di ra 2 băng; 2 - cá thể đồng hợp tử B/B có SNP nằm trong trình tự cắt của enzyme giới hạn nên không bị cắt, sản phẩm điện di ra 1 băng; 3 - cá thể dị hợp tử A/B chứa cả hai trình tự của bố và mẹ nên sản phẩm điện di ra 3 băng) (Nguồn: <http://ncbi.nlm.nih.gov>) [138]

Điện di sản phẩm cắt enzyme giới hạn trên gel agarose, so sánh sự đa hình của đoạn DNA thu được với bố, mẹ để xác định các dòng đồng hợp tử mang kiểu gen QTL9 của bố hoặc mẹ.

2.4.7. Phương pháp bố trí thí nghiệm và các chỉ tiêu theo dõi ngoài đồng ruộng

- Bố trí thí nghiệm đồng ruộng: Chọn ra 98 dòng F_2 đồng hợp tử bao gồm 49 dòng đồng hợp tử mang QTL9 của mẹ, 49 dòng đồng hợp tử mang QTL9 của bố và hai giống bố mẹ G6, G189 (làm đối chứng) sẽ được trồng trên đồng ruộng vụ xuân năm 2019. Thí nghiệm được thiết kế bằng phần mềm Microsoft Excel 2010 hàm Random. Bố trí thí nghiệm theo khối ngẫu nhiên hoàn chỉnh, mỗi khối là một dòng gồm 16 cây, trồng 4 hàng, các cây trồng trong cùng khối cách nhau 25 cm, khoảng cách giữa các khối là 40 cm (hình 2.5). Chi tiết sơ đồ bố trí thí nghiệm ngoài đồng ruộng trình bày trong phần phụ lục 4.



Hình 2.6. Mô hình bố trí các ô thí nghiệm ngoài đồng ruộng

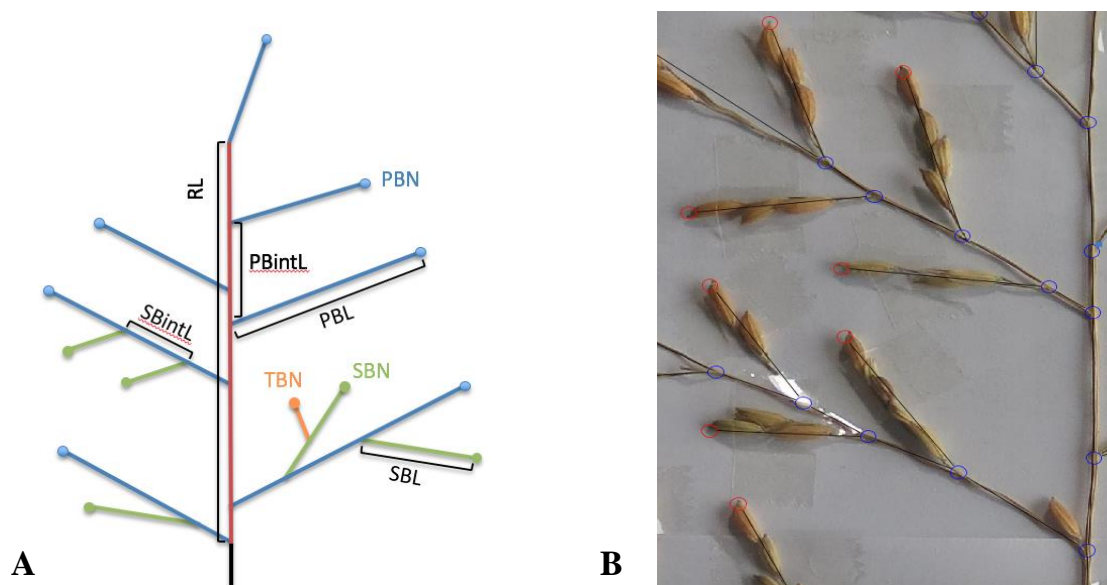
- Theo dõi thí nghiệm: Ba cây trồng ở giữa mỗi khối sẽ được theo dõi, đánh giá cấu trúc bông (hình 2.6) và một số tính trạng khác liên quan đến năng suất lúa: chiều cao cây, số ngày ra hoa bông đầu, số ngày ra hoa 50%, số nhánh/cây, số nhánh hữu hiệu/cây.

- Phương pháp đo đếm các chỉ tiêu liên quan đến năng suất lúa: Bằng phương pháp quan sát, đo đếm trực tiếp như sau:

+ Chiều cao cây (cm) được ghi lại ở giai đoạn trưởng thành và chiều cao của một cây được đo từ bề mặt đất cho đến đỉnh bông chính,

+ Số ngày ra hoa bông đầu tiên và số ngày ra hoa 50% (ngày) được tính từ ngày gieo hạt đến ngày có bông đầu tiên ra hoa và ngày bắt đầu ra hoa ở 50% số bông/cây,

+ Số nhánh/cây và số nhánh hữu hiệu/cây được xác định ở giai đoạn thu hoạch bằng cách đếm toàn bộ số nhánh/cây và số nhánh mang bông trên mỗi cây.



Hình 2.7. Cấu trúc bông lúa (A) và đưa vào phần mềm phân tích P-TRAP (B)

(RL: chiều dài bông, PBN: số gié cấp một/bông, SBN: số gié cấp hai/bông, TBN: số gié cấp ba/bông, SPN: Số hạt/bông, PBL: chiều dài gié cấp một, SBL: chiều dài gié cấp hai, PBintL: chiều dài giữa các gié cấp một, SBintL: chiều dài giữa các gié cấp hai). Nguồn: Tạ Kim Nhung, 2018.

- Phương pháp đo đếm các chỉ tiêu cấu trúc bông: Bằng phần mềm P-TRAP do nhóm Nghiên cứu cấu trúc bông lúa tại IRD-Montpellier AL-Tam và các cộng sự phát triển năm 2013 [12]: Sau 20 ngày ra hoa, 9 bông từ 3 cây ở giữa mỗi khối được thu thập, mỗi cây thu 3 bông. Bông lúa sau khi thu được dán cố định trên giấy trắng A₃, chụp ảnh bông và đưa vào phần mềm P-TRAP phân tích các chỉ tiêu bao gồm: Chiều dài bông (RL), số gié cấp một (PBN), chiều dài gié cấp một (PL), số gié cấp hai (SBN), chiều dài gié cấp hai (SBL), số gié cấp ba (TBN), khoảng cách giữa các gié cấp một (PBintL), khoảng cách giữa các gié cấp hai (SBintL) và số hạt/bông (SpN) (hình 2.6).

2.4.8. Phương pháp phân tích, xử lý số liệu

Phân tích thông kê các dữ liệu đo đếm được bằng cách sử dụng các chức năng khác nhau trong phần mềm R: Phân tích phương sai bằng hàm

ANOVA để đánh giá độ tin cậy của các dữ liệu, hàm *Shapiro test* (*Shapiro_test*, 2020) [96] được sử dụng để xác định các mô hình hóa có theo phân phối chuẩn (QQ-plot), sử dụng các gói công cụ *corrplot* (CRAN - Package *Corrplot*) [33], *ade4* (<https://cran.r-project.org/web/packages/ade4>) và *devtools* theo Wickham và các cộng sự (2021) [114] để phân tích mối tương quan kiểu hình giữa các tính trạng cấu trúc bông, phân tích hồi quy tuyến tính giữa hai tính trạng số gié cấp hai/bông (SBN) và số hạt/bông (SpN). Tất cả các phân tích thống kê, phân tích sự phân bố của quần thể và phân tích thành phần chính (PCA) của tập dữ liệu đều được thực hiện bằng phần mềm R theo (R Core Team (2019), Toparlan và các cộng sự (2020) [92], [108].

So sánh kiểu hình của các dòng F_3 đồng hợp tử mang kiểu gen QTL9 của bố hoặc mẹ với kiểu hình của bố, mẹ. Phân tích thống kê (Association test, variance test) bằng phần mềm R theo Nguyễn Văn Tuấn (2018) [8] và R Core Team, 2019 [92], từ đó có thể kết luận ảnh hưởng của QTL9 đến các tính trạng số gié cấp hai/bông và số hạt/bông ở các giống lúa bản địa.

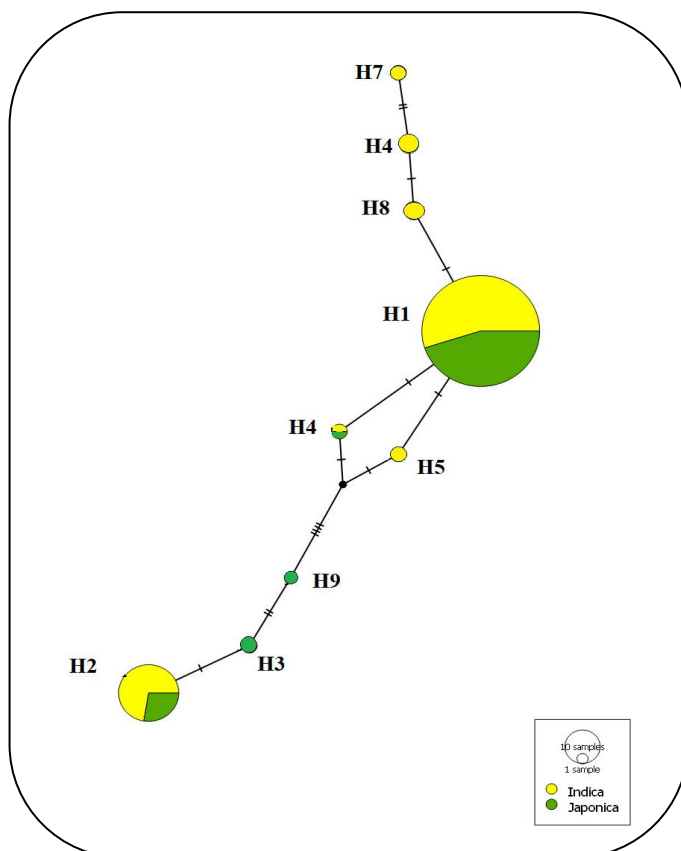
CHƯƠNG III

KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Chọn lọc các cặp lai bố mẹ và lai tạo quần thể F₁

3.1.1. Phân tích haplotype vùng QTL9 của tập đoàn lúa sử dụng trong nghiên cứu GWAS, chọn lọc các giống lúa bố mẹ làm vật liệu lai tạo

Kết quả phân tích liên kết của các SNPs trong vùng QTL9 của Tạ Kim Nhung và cộng sự (2018) [102] đã chỉ ra rằng tính trạng số gié cấp hai/bông và số hạt/bông cao liên kết với haplotype atataaat, còn tính trạng số gié cấp hai/bông và số hạt/bông thấp liên kết với haplotype gagagcgaa ở cùng vị trí (vị trí của SNPs) (Bảng phụ lục 1). Từ kết quả đó, tiến hành phân tích haplotype vùng QTL9 trên cả tập đoàn nghiên cứu GWAS gồm 155 giống lúa thu được 9 haplotype riêng biệt (hình 3.1, bảng 3.1).



Hình 3.1. Phân tích haplotype vùng QTL9

(Ghi chú: H1 - H9: 9 haplotype)

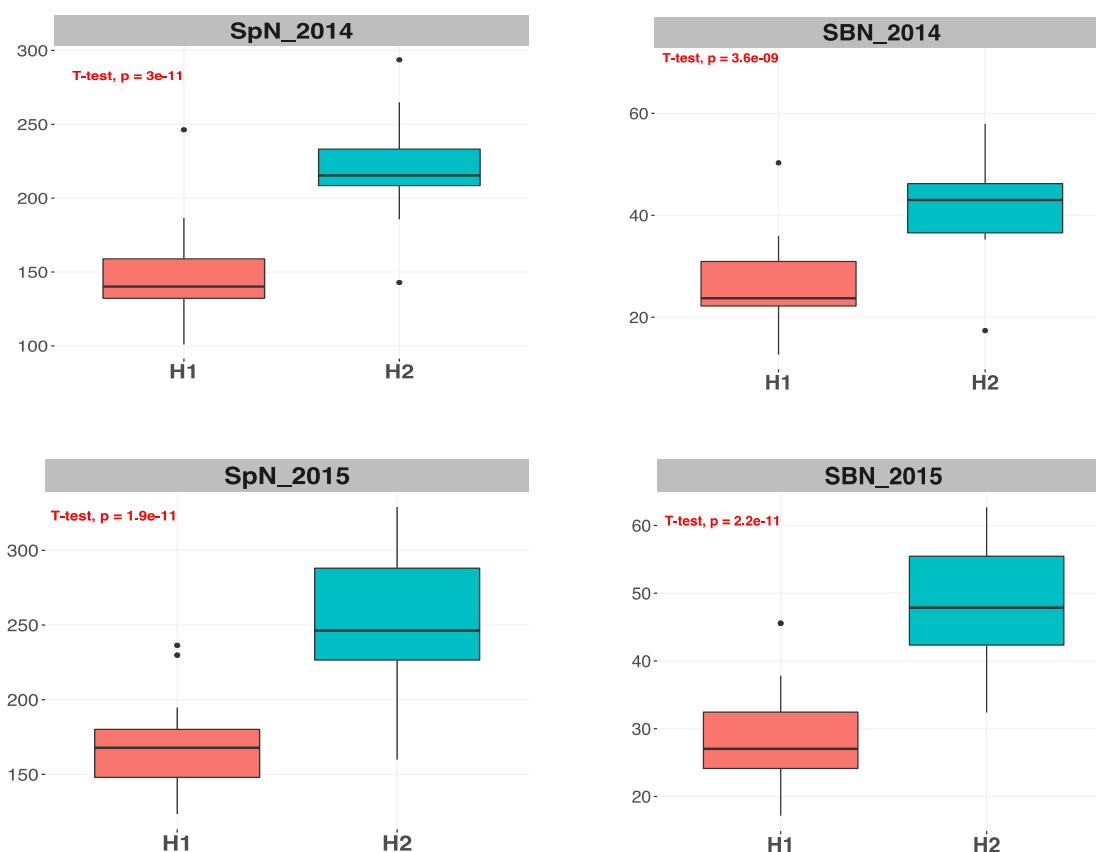
Bảng 3.1. Danh sách các giống lúa thuộc 9 haplotype và trình tự haplotype

STT	Haplotype	Ký hiệu giống		Trình tự haplotype
		Nhóm <i>Japonica</i>	Nhóm <i>Indica</i>	
1	H1	G100, G106, G107, G117, G124, G126, G128, G152, G154, G16, G177, G178, G187, G191, G193, G194, G195, G200, G202, G203, G204, G206, G210, G212, G214, G216, G217, G220, G221, G222, G223, G25, G26, G299, G38, G45, G46, G47, G48, G50, G61, G85, G86, G87, G88, G89, G90, G91, G92, G98	G1, G10, G102, G104, G11, G113, G115, G119, G12, G125, G129, G133, G138, G14, G140, G143, G146, G150, G165, G167, G168, G17, G170, G171, G173, G18, G19 , G192, G2, G201, G207, G21, G219, G22, G3, G4, G5, G51, G52, G53, G57, G58, G59, G6 , G62, G69, G7, G70, G72, G73, G74, G77, G79, G8, G9, G93, G94, G95, G96, G99, IR64	gagagcgaa
2	H2	G101, G130, G131, G134, G158, G80, G83, G84	G105, G110, G111, G132, G153, G155, G156, G160, G161, G169, G181, G182, G183, G186, G189 ,	atataaatt

STT	Haplotype	Ký hiệu giống		Trình tự haplotype
		Nhóm <i>Japonica</i>	Nhóm <i>Indica</i>	
			G190, G205, G208, G209, G63, G67	
3	H3	G103, G157		atataaata
4	H4		G109, G54, G78	gagagcggtt
5	H5		G172, G49	gtgagcgaa
6	H6	G179	G20	gagtgcgaa
7	H7		G37, G39	gtgagcatt
8	H8		G56, G64, G65	gagagcgat
9	H9	ACZ		atataagaa
	Tổng số	62 giống	93 giống	

Kết quả thu được ở hình 3.1 và bảng 3.1 cho thấy: Trong 9 haplotype đã phân tích có 3 haplotype H2, H3 và H9 liên kết với tổ hợp SNPs atataaatt/ atataaata/ atataagaa có số gié cấp hai/bông và số hạt/bông cao, 6 haplotype H1, H4, H5, H6, H7 và H8 liên kết với tổ hợp SNPs gagagcgaa/ gagagcggtt/ gtgagcgaa/ gagtgcgaa/ gtgagcatt/ gagagcgat có số gié cấp hai/bông và số hạt/bông thấp. Trong đó, hai haplotype khác biệt lớn nhất là H1 và H2 có khoảng cách di truyền lớn và có nhiều giống: Haplotype 1 (H1, gagagcgaa) có 111 giống (50 giống thuộc nhóm *Japonica*, 61 giống thuộc nhóm *Indica*), haplotype 2 (H2, atataaatt) có 29 giống (8 giống thuộc nhóm *Japonica*, 21 giống thuộc nhóm *Indica*). Các haplotype còn lại chỉ chứa từ 1 - 3 giống/haplotype (bảng 3.1).

Sau khi chọn lọc được hai haplotype H1, H2 có khoảng cách di truyền cách xa nhau, đa hình 9 SNP trong vùng QTL9 và có nhiều giống. Tiến hành phân tích kiểu hình cấu trúc bông của hai haplotype H1 và H2 để chọn lọc cây bố mẹ, kết quả phân tích thể hiện qua hình 3.2.



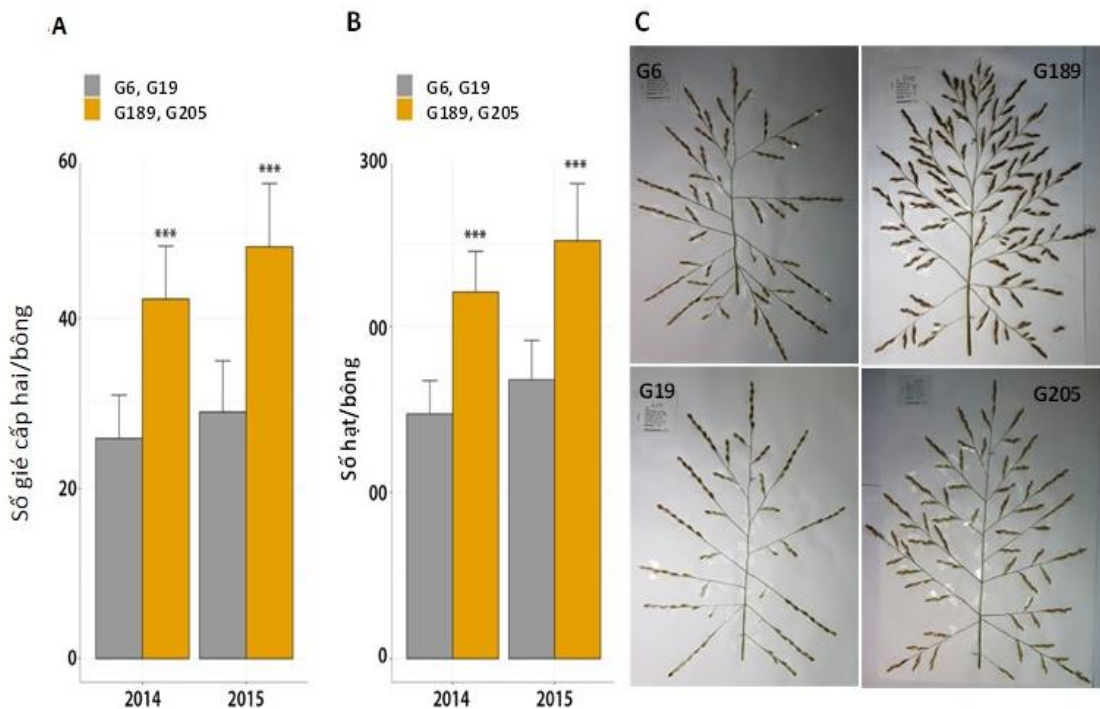
Hình 3.2. Phân tích cấu trúc bông của 2 haplotype đại diện (H1, H2)

(Ghi chú: SpN: Số hạt/bông; SBN: Số gié cấp hai/bông)

Hình 3.2 thể hiện kết quả phân tích hai tính trạng số gié thứ cấp/bông và số hạt/bông của các giống trong hai haplotype, sử dụng bộ dữ liệu cấu trúc bông thu được từ năm 2014 - 2015 của Tạ Kim Nhung và các cộng sự [102], kết quả phân tích cho thấy có sự khác biệt lớn có ý nghĩa thống kê ($p < 0,005$) trong cả hai năm 2014 và 2015. Các giống thuộc nhóm haplotype 1 (H1) có số lượng hạt/bông khoảng 150 hạt, số lượng gié cấp hai/bông khoảng 25 gié trong khi haplotype 2 (H2) có số lượng hạt trên bông khoảng 250 hạt, số gié cấp hai/bông khoảng 45 gié.

Mặt khác, so sánh thời gian sinh trưởng của các giống thuộc hai haplotype từ bộ dữ liệu bảng phục lục 1 cho thấy, H1 có 28 giống thuộc nhóm thời gian sinh trưởng sớm (< 85 ngày), 48 giống có thời gian sinh trưởng

trung bình (từ 85 ngày đến 105 ngày) và 35 giống có thời gian sinh trưởng muộn (trên 135 ngày). Trong khi đó, H2 có 5 giống thuộc nhóm thời gian sinh trưởng sớm, 9 giống có thời gian sinh trưởng trung bình và 15 giống có thời gian sinh trưởng muộn. Từ đó, bốn giống lúa thuộc hai haplotype và có kiểu hình cấu trúc bông khác biệt đã được chọn ra (hình 3.3).



Hình 3.3. Hình thái cấu trúc bông của 4 giống lúa sử dụng làm bố, mẹ để tạo quần thể lai F₁

Ghi chú: A: số gié cấp hai/bông của 2 nhóm lúa bông nhỏ G6, G19 (màu ghi) và bông to G189, G205 (màu vàng); B: số hạt/bông của 2 nhóm lúa bông nhỏ G6, G19 (màu ghi) và bông to G189, G205 (màu vàng); C: Ảnh bông lúa của 4 giống G6, G19, G189, G205.

Kết quả hình 3.3 cho thấy, hai giống G6 và G19 thuộc H1 (gagagcga) có kiểu hình bông nhỏ, số gié cấp hai/bông dao động từ 23 - 34 gié đối với giống G6 và 24 - 31 gié đối với giống G19, số hạt/bông của giống G6 từ 130 - 187 hạt trong khi giống G19 đạt 133 - 177 hạt. Hai giống có kiểu hình bông to G189 và G205 thuộc H2 (atataaatt) có số gié cấp hai/bông khoảng 45 - 48 gié

và số hạt/bông từ 220 - 250 hạt. Cả bốn giống lúa được chọn đều cùng loài (*Indica*) và có thời gian sinh trưởng tương đương nhau, dao động từ 94,5 - 108 ngày.

Từ kết quả phân tích trên, bốn giống lúa G6, G19, G189, G205 được chọn làm bố mẹ để tạo quần thể lai F_1 . Trong đó, hai giống G6, G19 thuộc haplotype 1, cấu trúc bông nhỏ được chọn làm cây mẹ; hai giống G189, G205 thuộc haplotype 2, cấu trúc bông to làm cây bố.

Mặc dù có nhiều QTL/gen chi phối các tính trạng số lượng khác nhau đã được báo cáo ở lúa. Nhưng việc tổ hợp các nhóm haplotype ưu việt của chúng để tạo ra giống lý tưởng vẫn gặp nhiều khó khăn. Theo công bố gần đây trên tạp chí Công nghệ sinh học thực vật của Abbai và cộng sự (2019) [9] đã phân tích haplotype của 120 gen ảnh hưởng đến năng suất và chất lượng hạt lúa trong bộ 3K genom lúa, trong đó xác định được 21 gen liên kết chặt chẽ chi phối năng suất và chất lượng của 10 tính trạng. Dựa trên dữ liệu kiểu hình của quần thể lúa thu được 15 haplotype (H) với tần số haplotype (HF) từ 1,65 - 65,84%, tần số thấp nhất là H6 (1,65%) cho ra hoa muộn, tần số cao nhất là H4 (65,84%) cho hạt mảnh. Phân tích haplotype có thể giải thích cơ sở di truyền có thể có cho tính ưu việt của các giống đã chọn như các nghiên cứu của Abbai và cộng sự (2019) [9] và Bhat và cộng sự (2021) [27].

Bằng cách sử dụng bộ dữ liệu giải mã lại toàn bộ bộ gen quy mô lớn kết hợp với phân tích kiểu hình haplotype, Abbai và cộng sự (2019) [9] đã xác định được các haplotype hữu ích cho việc nhân giống lúa trong tương lai và Sinha và cộng sự [98] đã làm theo cách tiếp cận tương tự ở cây đậu triều (*Pigeon pea*). Bên cạnh đó, việc tích hợp bộ gen để xác định các tái tổ hợp được tạo ra bằng cách lai giữa các cặp bố mẹ tương phản sẽ hỗ trợ rất nhiều trong việc giải quyết sự phức tạp của các tính trạng số lượng của Jensen và cộng sự (2020) [55].

Như vậy, việc phân tích haplotype để chọn lọc các giống lúa bố mẹ phục vụ lai tạo quần thể F_1 như trong nghiên cứu của luận án là phù hợp với mục tiêu và các nghiên cứu trước đây đã công bố.

3.1.2. Tạo quần thể lai F_1 và chọn lọc cây F_1 bằng chỉ thị phân tử SSR

3.1.2.1. Tạo quần thể F_1

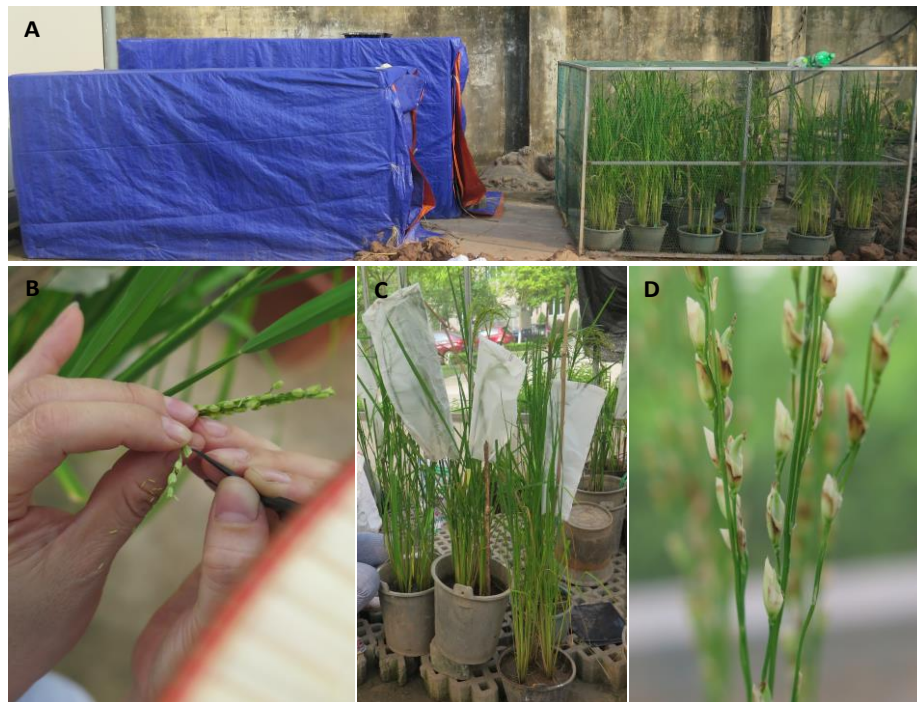
Bốn quần thể F_1 được tạo ra từ bốn cặp lai giữa các giống bố mẹ có cấu trúc bông tương phản (bông to và bông nhỏ) đã được chọn lọc ở trên. Bốn cặp lai được tiến hành từ bốn giống lúa bố mẹ G6, G19 và G189, G205 như sau:

Bảng 3.2. Các cặp lai được tiến hành từ 4 giống bố mẹ (G6, G19, G189, G205)

♀ \ ♂	G189 (bông to)	G205 (bông to)
G6 (bông nhỏ)	G6 x G189	G6 x G205
G19 (bông nhỏ)	G19 x G189	G19 x G205

Các giống G6, G205 có thời gian ra hoa muộn hơn giống G19, G189 nên được trồng trước 1 - 2 tuần. Ngoài ra, các giống lúa đều mẫn cảm với quang chu kỳ, chỉ ra hoa vào ngày ngắn nên chu kỳ quang đã được xử lý sau 2 tháng gieo trồng cho đến khi cây lúa trở bông, đảm bảo 9 giờ chiếu sáng/ngày (từ 8 giờ sáng đến 17 giờ chiều) nhằm mục đích điều khiển sự ra hoa đồng đều giữa các giống. Kết quả là cả bốn giống lúa đều ra hoa gần như cùng thời điểm. Năm ngày sau khi lai, hạt lai F_1 đã xuất hiện trên bông ở tất cả các cây mẹ (G6, G19). (hình 3.4).

Kết quả quá trình lai tạo thu được 328 hạt F_1 của cặp lai G6 và G189; 261 hạt F_1 của cặp lai G6 và G205; 191 hạt F_1 của cặp lai G19 và G189; 152 hạt F_1 của cặp lai G19 và G205. Các cây F_1 sẽ được chọn lọc bằng chỉ thị phân tử SSR.



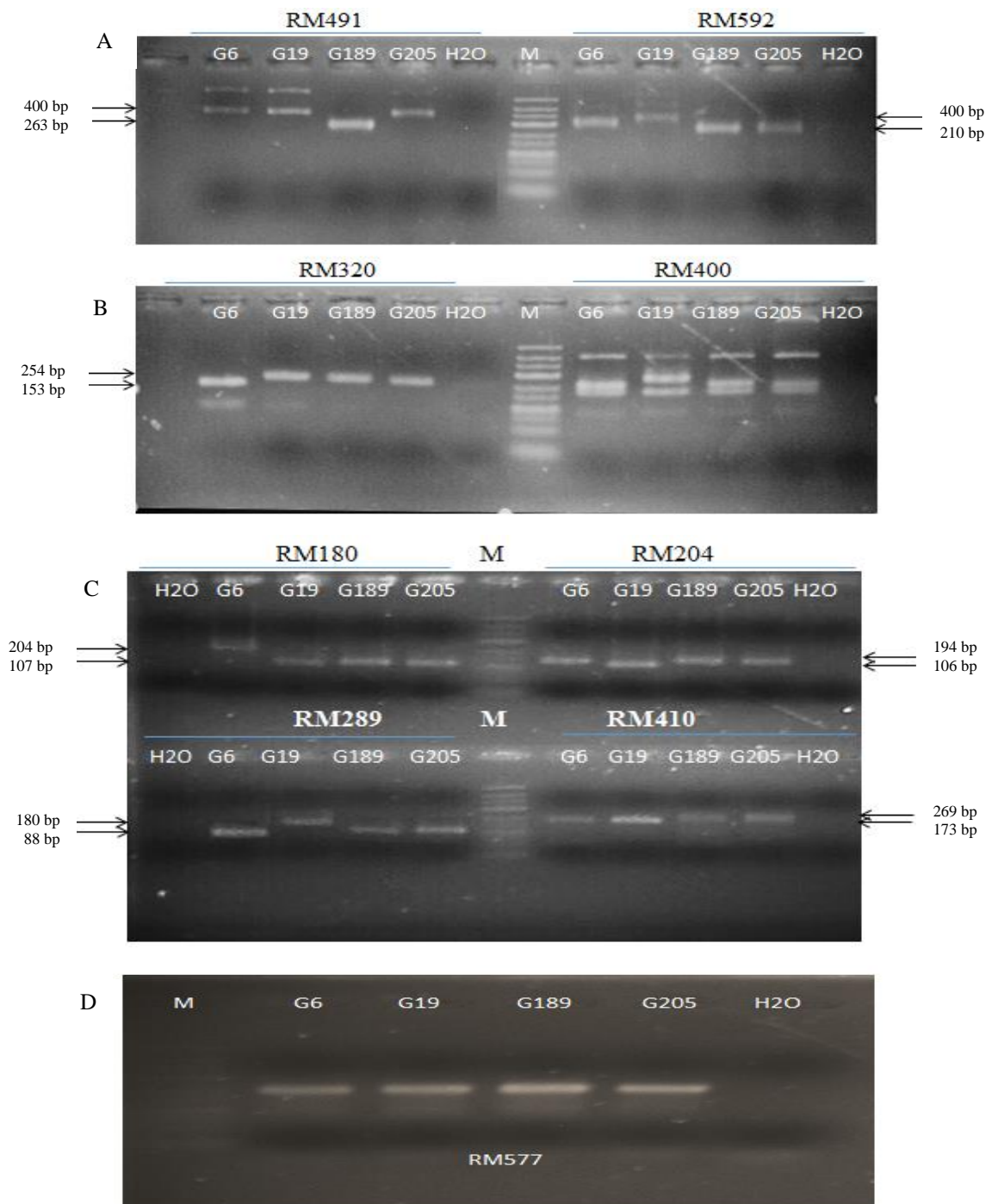
Hình 3.4. Quá trình xử lý chu kỳ quang và lai giữa các giống lúa bố mẹ để tạo quần thể F_1 (tại Viện Di truyền Nông nghiệp năm 2017)

(Ghi chú: A - Buồng xử lý tối và sáng theo chu kỳ quang 9 giờ chiếu sáng và 15 giờ trong bóng tối cho tất cả các giống lúa bố mẹ. B - Loại bỏ bao phấn trên cây mẹ (G6, G19). C - Bảo vệ bông lúa của cây mẹ sau khi lai bằng bao giấy. D - Hạt F_1 sau 5 ngày lai)

3.1.2.2. Chọn lọc chỉ thị phân tử SSR cho đa hình giữa hai giống lúa bố mẹ

Hơn 2000 chỉ thị SSR được công bố cho đa hình ở các giống lúa *Indica* (<http://www.gramene.org/>) [131]. Mười hai chỉ thị SSR được chọn lọc để đánh giá kiểu gen của các quần thể F_1 (Bảng 2.2), các chỉ thị nằm trên các nhiễm sắc thể khác nhau và có sự đa hình lớn, trình tự lặp khác nhau và cho phép quan sát được trên gel agarose.

Trước khi sử dụng chỉ thị SSR để chọn lọc các cây F_1 thì 12 chỉ thị SSR (Bảng 2.2) được đánh giá trên các cây bố mẹ nhằm chọn lọc chỉ thị cho đa hình giữa hai giống lúa bố mẹ. Kết quả khảo sát sự đa hình của các chỉ thị phân tử SSR giữa các giống lúa bố mẹ thể hiện qua hình 3.5.



Hình 3.5. Kết quả khảo sát sự đa hình của các chỉ thị phân tử SSR ở các giống bố mẹ (G6, G19, G189, G205)

(A: khảo sát sự đa hình của hai chỉ thị phân tử RM491 và RM592; B: khảo sát sự đa hình của hai chỉ thị phân tử RM320 và RM400; C: khảo sát sự đa hình của bốn chỉ thị phân tử RM180, RM204, RM289, RM410; D: khảo sát sự đa hình của chỉ thị RM577; M: Thang chuẩn DNA 100 bp)

Phân tích kết quả PCR (hình 3.5) cho thấy bảy chỉ thị SSR cho sự đa hình giữa các giống bố mẹ. Trong đó hai chỉ thị RM320 và RM180 cho sự đa hình giữa giống mẹ G6 và cả hai giống bố G189, G205 (hình 3.5B, 3.5C), do đó hai chỉ thị RM320 và RM180 có thể được sử dụng để phân tích quần thể F_1 của hai cặp lai G6 x G189 và G6 x G205.

Hai chỉ thị RM204 và RM289 cho đa hình giữa giống mẹ G19 và hai giống bố G189, G205 nên có thể sử dụng để phân tích quần thể lai F_1 của hai cặp G19 x G189 cũng như G19 x G205 (hình 3.5C).

Chỉ thị phân tử RM491 mặc dù không đặc hiệu ở hai giống mẹ G6 và G19 nhưng vẫn cho sự đa hình giữa hai giống so với giống G189 (hình 3.5A) nên có thể sử dụng để chọn lọc cây F_1 của hai cặp lai G6 x G189 và G19 x G189.

Hai chỉ thị phân tử RM592 (hình 3.5A) và RM410 (hình 3.5C) cho sự đa hình giữa hai nhóm giống bố, mẹ nên có thể sử dụng để xác định cây F_1 của cả bốn cặp lai.

Năm chỉ thị RM151, RM400, RM532, RM535 và RM577 (hình 3.5D) không cho đa hình giữa các giống bố mẹ nên không được sử dụng để chọn lọc các cây F_1 .

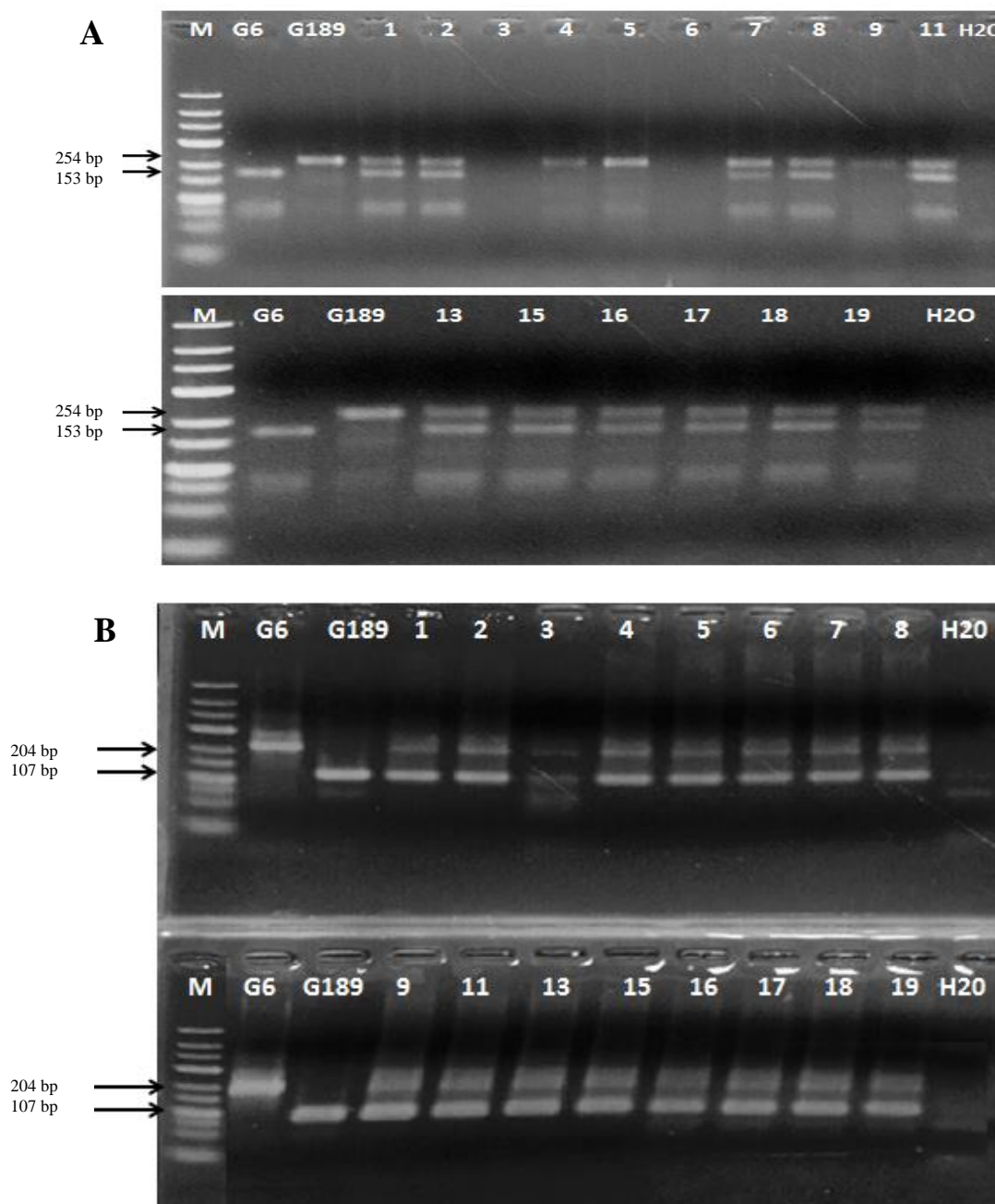
3.1.2.3. Chọn lọc cây F_1 bằng chỉ thị phân tử SSR

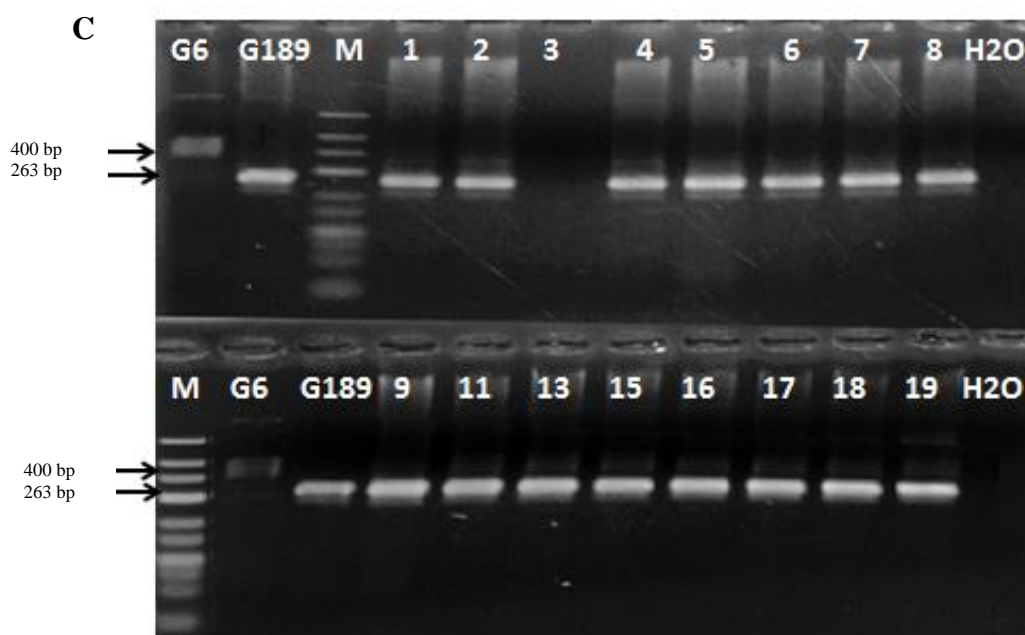
a, Chọn lọc cây F_1 của cặp lai G6 x G189

Hai mươi cây F_1 của cặp lai G6 x G189 được phân tích với ba chỉ thị phân tử RM320, RM180, RM491. Cây được xác định là cây lai nếu mang cả hai đoạn DNA tương ứng với chỉ thị ở cây bố và cây mẹ. Kết quả thể hiện hình 3.6.

Phân tích hình ảnh điện di của chỉ thị RM320 (hình 3.6A) đã xác định được 11 cây F_1 (1, 2, 7, 8, 11, 14, 15, 16, 17, 18, 19) mang cả hai đoạn DNA của mẹ và bố có chiều dài lần lượt là 153 pb và 254 pb. Tương tự, đối với chỉ thị RM180 (hình 3.6B), số cây F_1 mang cả hai đoạn DNA có kích thước 107

bp và 204 bp tương ứng với đoạn DNA ở cây mẹ và cây bố là 16 cây (1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 15, 16, 17, 18, 19). Đối với chỉ thị RM491 (hình 3.6C) số cây F₁ được xác định là 10 cây (1, 2, 4, 11, 13, 15, 16, 17, 18, 19) mang cả hai đoạn DNA có kích thước 263 bp và 400 bp tương ứng với đoạn DNA ở cây mẹ và cây bố.





Hình 3.6. Kết quả chọn lọc cây F_1 của cặp lai G6 x G189 với ba chỉ thị phân tử RM320 (A), RM180 (B) và RM491 (C).

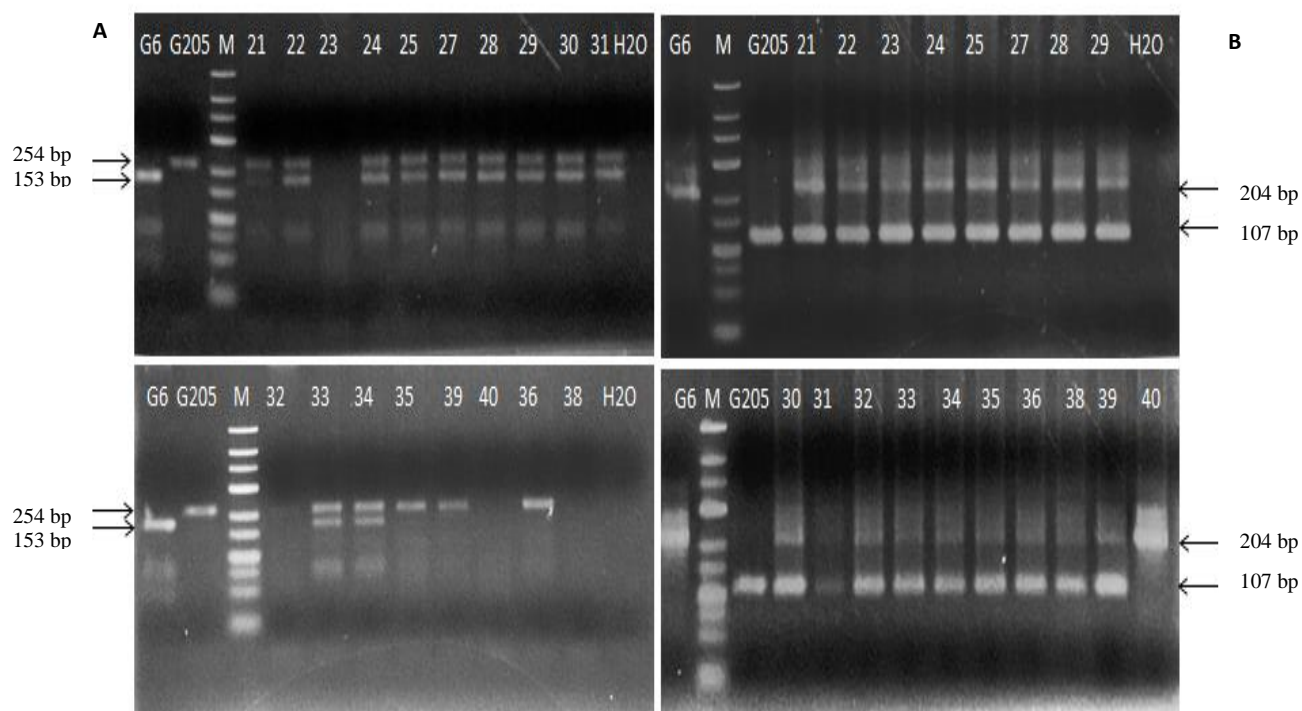
M: thang chuẩn DNA 100 bp (Promega)

Tổng hợp kết quả phân tích với ba chỉ thị phân tử, 12 cây F_1 được xác định là cây lai, khẳng định bởi ít nhất với hai chỉ thị phân tử: 1, 2, 4, 7, 8, 11, 13, 15, 16, 17, 18, 19.

b, Chọn lọc cây F_1 của cặp lai G6 x G205

Hai mươi cây F_1 của cặp lai G6 x G205 được phân tích bằng hai chỉ thị phân tử RM320 và RM180 (hình 3.7).

Kết quả hình 3.7 cho thấy: Với chỉ thị RM320 (hình 3.7A) đã xác định được 11 cây F_1 (21, 22, 24, 25, 27, 28, 29, 30, 31, 33, 34) mang cả hai đoạn DNA có chiều dài tương ứng ở giống mẹ G6 (153 bp) và ở giống bố G205 (254 bp). Chỉ thị RM180 (hình 3.7B) đã xác định được 16 cây F_1 (21, 22, 23, 24, 25, 27, 28, 29, 30, 32, 33, 34, 35, 36, 38, 39) mang cả hai lô-cut của bố (107 bp) và mẹ (204 bp). Các cây còn lại chỉ mang một băng DNA không được xác định là cây lai. Tổng hợp kết quả phân tích với cả hai chỉ thị phân tử, 11 cây F_1 đã được chọn lọc, bao gồm 21, 22, 24, 25, 26, 27, 28, 29, 30, 33, 34.

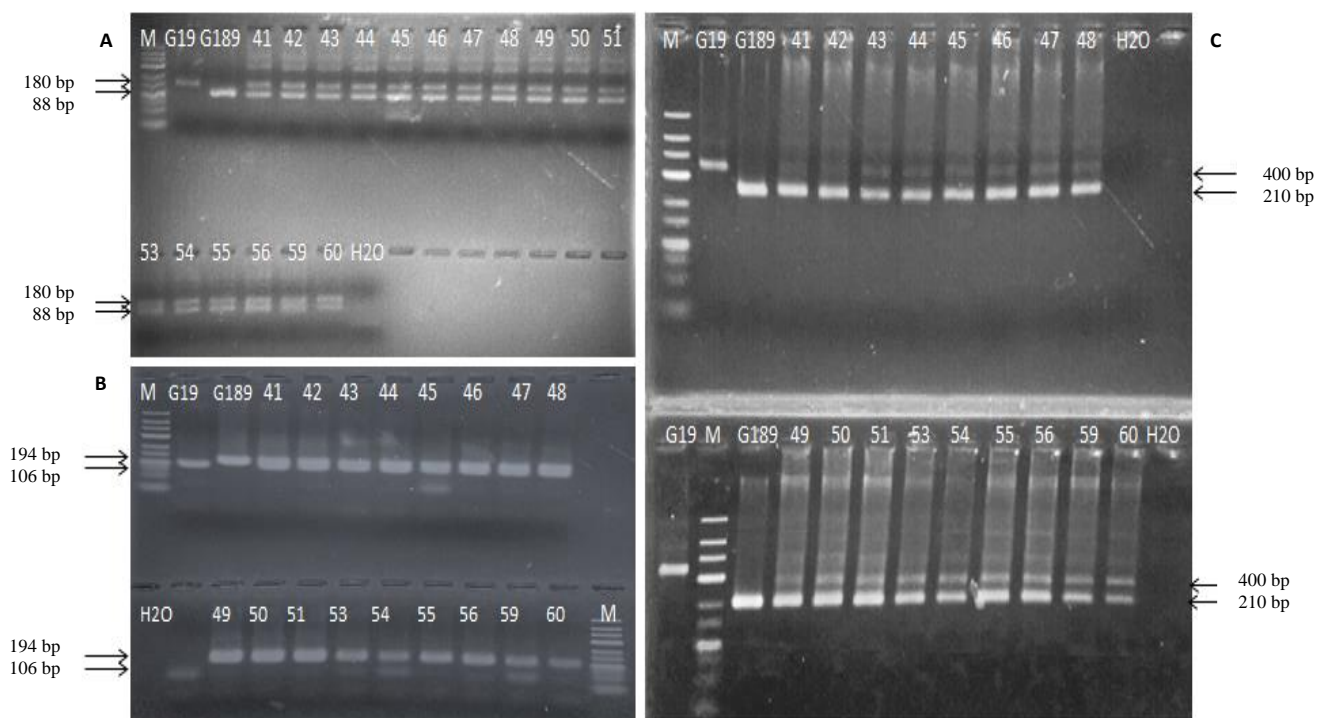


Hình 3.7. Kết quả chọn lọc cây F_1 của cặp lai G6 x G205 với hai chỉ thị phân tử RM320 (A), RM180 (B). M: thang chuẩn DNA 100 bp (Promega).

c, Chọn lọc cây F_1 của cặp lai G19 x G189 và cặp lai G19 x G205

Hai mươi cây F_1 của mỗi cặp lai được phân tích bằng ba chỉ thị phân tử RM289, RM592 và RM204. Kết quả thể hiện hình 3.8.

Hình 3.8 cho thấy: Đối với cặp lai G19 x G189, sử dụng chỉ thị RM289 cho thấy 17 cây F_1 (41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 53, 54, 55, 56, 59, 60) đều là cây lai vì mang cả 2 băng DNA có chiều dài tương ứng với băng đa hình ở giống lúa bố mẹ (88 bp và 180 bp) (hình 3.8A). Tương tự như vậy chỉ thị RM592 cũng cho kết quả 17 cây đều là cây lai (hình 3.8B). Với chỉ thị RM204 có 11 cây lai (hình 3.8C). Như vậy từ cặp lai G19 x G189 đã chọn ra được 17 cây F_1 lai được khẳng định ít nhất bởi hai chỉ thị.

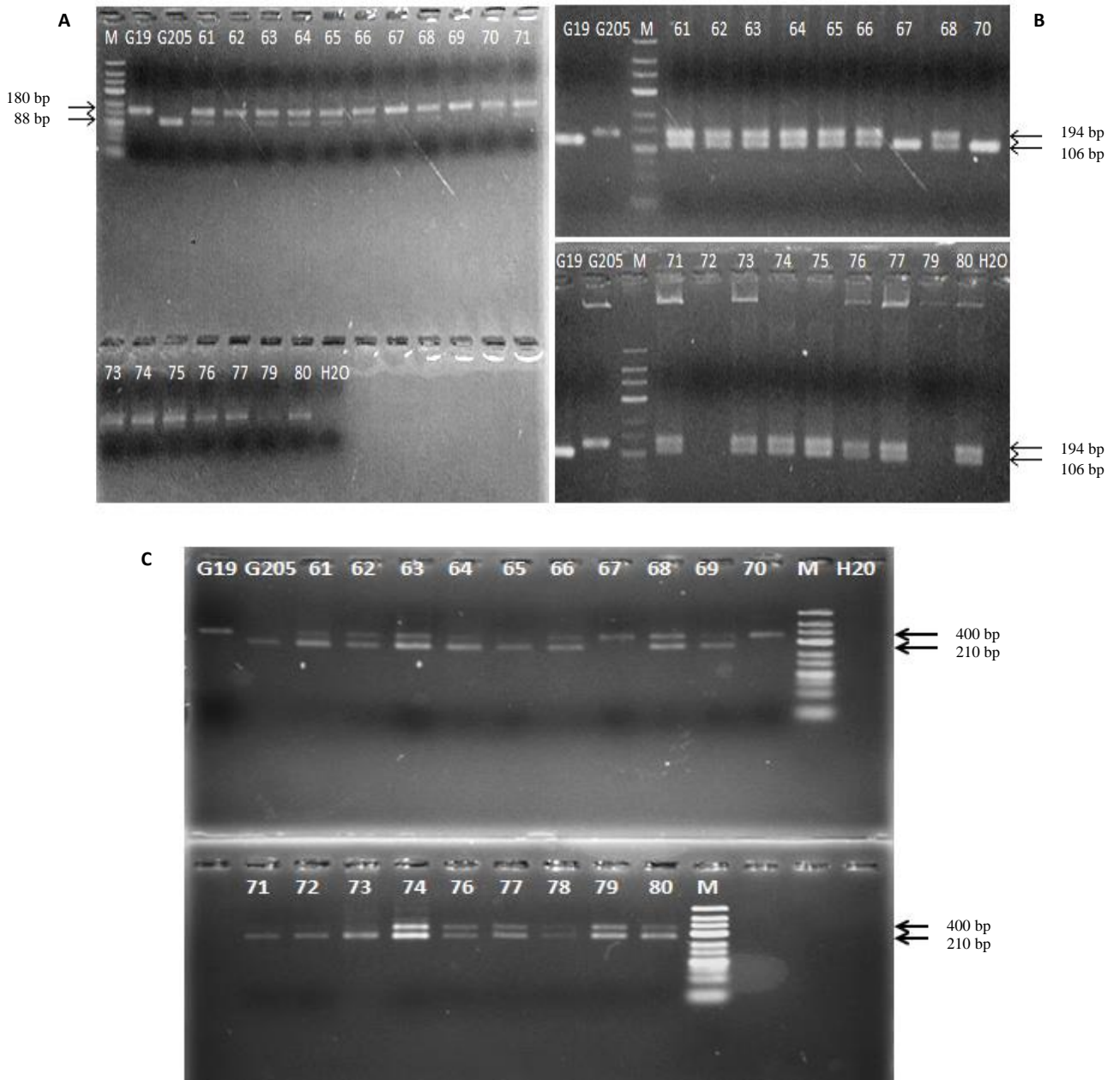


Hình 3.8. Kết quả chọn lọc cây F₁ của cặp lai G19 x G189 với ba chỉ thị phân tử RM289 (A), RM592 (B) và RM204 (C)

M: thang chuẩn DNA 100 bp (Promega)

Tiếp tục kiểm tra cặp G19 x G205 bằng 3 chỉ thị RM289, RM592 và RM204 cho kết quả thể hiện hình 3.9.

Kết quả hình 3.9 cho thấy, với chỉ thị RM289, các cây 61, 62, 63, 64, 65, 66 được xác định là cây lai vì mang cả hai đoạn đa hình của các giống lúa bố mẹ (88 bp và 180 bp) (hình 3.9A). Chỉ thị RM592 cho kết quả các cây 61, 62, 63, 64, 65, 66, 68, 74, 76, 77, 78, 79, 80 là các cây lai mang cả hai lô-cut của bố và mẹ (210 bp và 400 bp) (hình 3.9B). Đối với chỉ thị RM204, cũng xác định được 14 cây lai mang cả hai lô-cut của bố và mẹ (106 bp và 194 bp) (hình 3.9C). Tổng kết kết quả của cả ba chỉ thị phân tử SSR, chọn được 11 cây lai bao gồm các cây 61, 62, 63, 64, 65, 66, 68, 74, 76, 77, 80.



Hình 3.9. Kết quả chọn lọc cây F₁ của cặp lai G19 x G205 với ba chỉ thị phân tử RM289 (A), RM592 (B) và RM204 (C)

M: thang chuẩn DNA 100 bp (Promega)

Như vậy, kiểm tra 20 cây F₁ ở mỗi cặp lai bằng chỉ thị phân tử SSR đã xác định được: 12 cây lai F₁ của cặp G6 x G189 nhờ 3 chỉ thị RM320, RM180, RM491; 11 cây lai F₁ của cặp G6 x G205 nhờ 2 chỉ thị RM320, RM180; 17

cây lai F_1 của cặp G19 x G189 nhờ 3 chi thị RM289, RM592, RM204; 11 cây lai F_1 của cặp G19 x G205 nhờ ba chi thị RM289, RM592, RM204.

Số hạt F_1 của cặp lại G6 x G189 thu được nhiều nhất nên được chọn để phát triển quần thể F_2 tự thụ bao gồm ít nhất 200 cá thể.

3.2. Phát triển chi thị phân tử CAPS để chọn lọc các cây F_2 mang QTL9 đồng hợp thuộc hai haplotype bố hoặc mẹ

3.2.1. Xác định các SNPs trong vùng QTL9 của hai giống bố mẹ để phát triển chi thị phân tử CAPS

Ứng dụng công nghệ chụp gen kết hợp với giải trình tự thế hệ mới Illumina cho phép bắt và xác định trình tự của 72 trong số 77 gen (bao gồm cả vùng promoter và terminator) trong vùng QTL9 và tạo ra tổng cộng 1.830.925 trình tự đọc. Kích thước trung bình của các lần đọc dao động là 350 bp. Một bước phân tách với fastq-multx cho phép tách các lần đọc thuộc về từng mẫu. Dữ liệu kém chất lượng, adaptor và index được loại bỏ, sử dụng công cụ Cutadapt. Ở giai đoạn phân tích này, 1,8 triệu lượt đọc đã được truy xuất và có thể được sử dụng cho các giai đoạn ánh xạ. Fastqc được sử dụng để kiểm định lại chất lượng của các lần đọc sau các bước làm sạch.

Dữ liệu trình tự đọc thô được đánh giá, kiểm tra chất lượng và nhận diện các lỗi trong dữ liệu bằng công cụ Fastqc. Sau đó dữ liệu được giống hàng với bộ gen tham chiếu Nipponbare và loại bỏ vị trí phân tử trùng lặp, sử dụng phần mềm BWA-MEM. BWA (Burrows-Wheeler Alignment Tool) là một chương trình phần mềm liên kết trình tự các gen nhỏ khác nhau với một bộ gen tham khảo lớn.

Chương trình này bao gồm 3 thuật toán BWA-backtrack, BWA-SW và BWA-MEM. Thuật toán đầu tiên BWA-backtrack được thiết kế cho việc đọc chuỗi trình tự Illumina có kích thước 100 bp trở xuống, trong khi hai thuật toán kia dùng cho các trình tự có khả năng đọc cao hơn, dao động từ 70 bp

đến 1 Mbp. BWA-MEM và BWA-SW chia sẻ các chức năng tương tự nhau, ví dụ như hỗ trợ khả năng đọc cao và sắp xếp các trình tự. Tuy nhiên, BWA-MEM là chương trình mới nhất và được khuyến cáo dùng cho các kết quả có yêu cầu chất lượng, độ chính xác cao và nhanh hơn.

Thêm vào đó, BWA-MEM còn có hiệu suất tốt hơn so với BWA-backtrack trong khoảng đọc 70 - 100 bp. Đối với tất cả các thuật toán của BWA, việc cần thiết đầu tiên là phải cấu trúc được FM-index cho các gen tham khảo (sử dụng lệnh index). Các thuật toán sắp xếp được thực hiện theo lệnh “aln/samse/sample”, “bwasw” đối với BWA-SW và “mem” đối với BWA-MEM. Kết quả giống hàng dữ liệu được sàng lọc bằng phần mềm SAMtools để đảm bảo các đoạn trình tự đọc đã được giống hàng chính xác với bộ gen tham chiếu và loại bỏ những sai sót. Picard là bộ công cụ được xây dựng trên nền tảng Java nhằm thao tác trên tập tin định dạng SAM, BAM. Picard MarkDuplicates sẽ kiểm tra việc sắp xếp dữ liệu trong tập SAM và BAM qua đó cung cấp vị trí các phân tử trùng lặp.

Kết quả bảng 3.3 cho thấy hơn 98,5% dữ liệu được giống hàng thành công với trình tự tham chiếu Nipponbare (1.802.839 trình tự giống hàng thành công/ 1.830.925 trình tự đọc). Sau khi sử dụng Picard để loại bỏ phân tử trùng lặp, hơn 98% số đoạn trình tự được giữ lại, trong đó 68,67 - 70,1% dữ liệu được ánh xạ vào vùng mục tiêu QTL9 (Bảng 3.3).

Bảng 3.3. Kết quả giống hàng

Tên mẫu	Số đoạn trình tự giống hàng thành công	Số đoạn trình tự giống hàng thành công sau khi loại bỏ phân tử trùng lặp	Số đoạn trình tự được ánh xạ vào vùng mục tiêu QTL9
G6	1.023.987	1.006.256 (98,26%)	714.128 (70,1%)
G189	778.852	765.081 (98,23%)	512.416 (68,67%)

Các biến thể (SNPs, INDELs) được phát hiện bằng công cụ Haplotypecaller. HaplotypeCaller có khả năng gọi SNP và indels đồng thời thông qua tổ hợp de-novo cục bộ của các haplotype trong một vùng hoạt động. Nói cách khác, bất cứ khi nào chương trình gặp một vùng có dấu hiệu thay đổi, chương trình sẽ loại bỏ thông tin ánh xạ hiện có và tập hợp lại hoàn toàn các lần đọc trong vùng đó. Điều này cho phép HaplotypeCaller chính xác hơn khi gọi các vùng theo truyền thống khó gọi, chẳng hạn như khi chúng chứa các loại biến thể gần nhau.

Kết quả thu được lần lượt 4524 và 4522 biến thể cho hai giống G6 và G189. Sau đó những biến thể có tần suất allele thấp hơn 5% (Minor Allele Frequency - MAF < 5%) và những biến thể dị hợp hoặc thiếu dữ liệu bị loại bỏ bằng công cụ SNIPlay, 1178 biến thể cùng vị trí được giữ lại cho cả hai giống.

So sánh với bộ dữ liệu di truyền DArt (Diversity Arrays Technology) của Phùng Kim Nhung và các cộng sự năm 2014 [89], bộ dữ liệu GBS (Genotyping By Sequencing) Phùng Kim Nhung và các cộng sự năm 2016 [88] và dữ liệu kiểu hình cấu trúc bông của hai haplotype của Tạ Kim Nhung và cộng sự năm 2018 [102], những biến thể không có mặt trong các bộ dữ liệu này bị loại bỏ. Bộ dữ liệu cuối cùng thu được giữa hai haplotype có 1002 biến thể đồng hợp trong đó có 827 SNPs và 175 INDELs. Trong đó, 117 biến thể nằm trong vùng mã hóa exon, 391 biến thể nằm trong vùng không mã hóa intron, 320 biến thể nằm trong vùng promoter, 55 biến thể nằm trong vùng UTR-3', 7 biến thể nằm trong vùng UTR-5' và 112 biến thể nằm trong vùng giao thoa giữa các gen (Bảng 3.4).

Bảng 3.4. Kết quả tìm và sàng lọc biến thể

Tên biến thể	Tổng số	Vị trí					Vùng giao thoa giữa các gen
		Exon	Intron	Promoter	UTR-3'	UTR-5'	
SNP	827	107	325	256	38	6	95
INDEL	175	10	66	64	17	1	17
Tổng số	1002	117	391	320	55	7	112

Sau khi sàng lọc và phân chia các biến thể đồng hợp. Tiến hành chú giải biến thể được thực hiện bằng phần mềm SnpE. Phần mềm SnpE sử dụng để phân chia các biến thể thành các nhóm theo mức độ ảnh hưởng chức năng của biến thể. Đây là công cụ chú thích và dự báo ảnh hưởng của các biến thể đến cấu trúc protein như thay đổi axit amin, làm lệch khung đọc, thêm bộ ba mã hóa, mất bộ ba mã hóa...(Bảng 3.5).

Dữ liệu đầu vào của công cụ này là các biến thể được dự đoán (SNPs, chèn, xóa và MNPs), là kết quả của giải trình tự, và có định dạng VCF (Variant Call Format). Trong dữ liệu đầu ra, SnpE sẽ phân tích các biến thể đầu vào để chú giải và tính toán các tác động mà các biến thể có thể tạo ra trên gen. SnpE đưa ra các kết quả như sau: kiểu gen và các điểm bị ảnh hưởng bởi biến thể; vị trí của các biến thể; làm thế nào mà các biến thể ảnh hưởng đến quá trình tổng hợp protein; so sánh với các dữ liệu khác để tìm các biến thể đã biết (Bảng 3.6, 3.7, 3.8, 3.9 và 3.10).

Các bước chú giải biến thể và chuyển hóa các file được thực hiện bằng công cụ Galaxy của platforme Southgreen, sử dụng SNIPlay3, một ứng dụng trực tuyến để khám phá và phân tích quy mô lớn về các biến thể gen của Dereeper và các cộng sự năm 2011 [37]. Các biến thể nằm trong bộ ba mã hóa nhưng không dẫn đến làm thay đổi axit amin được gọi là biến thể đồng nghĩa, ngược lại biến thể nằm trong bộ ba mã hóa dẫn đến làm thay đổi axit

amin được gọi là biến thể sai nghĩa. Các biến thể thêm bớt (INDELs) làm lệch khung đọc chuẩn hoặc mất bộ ba mã hóa. Kết quả thu được 6 nhóm biến thể (Bảng 3.5).

Bảng 3.5. Phân loại biến thể trong vùng exon

Tên biến thể	SNP	INDEL
Hiệu ứng		
Biến thể đồng nghĩa	38	
Biến thể sai nghĩa	62	
Đột biến lệch khung đọc	0	8
Thêm bộ ba mã mở đầu	1	
Thêm bộ ba mã kết thúc	6	
Mất bộ ba mã hóa	0	2
Tổng cộng	107	10

Bảng 3.5 cho thấy: Trong 117 biến thể tạo ra nằm trong vùng exon có 107 biến thể SNP, 10 biến thể chèn xóa INDEL. Trong 107 biến thể SNP có 38 biến thể đồng nghĩa, 62 biến thể sai nghĩa, 1 biến thể thêm bộ mã mở đầu, 6 biến thể thêm bộ ba kết thúc và không có biến thể đột biến lệch khung đọc hay mất bộ ba mã hóa. Trong khi đó, biến thể INDEL có 8 biến thể đột biến lệch khung đọc và 2 biến thể mất bộ ba mã hóa.

Vị trí và những hiệu ứng cụ thể của 117 biến thể nằm trong vùng exon giữa hai haplotype được chú giải chi tiết, kết quả thể hiện qua các bảng 3.6, 3.7, 3.8, 3.9 và 3.10 sau.

Bảng 3.6. Chú giải biến thể đồng nghĩa giữa hai haplotype

Số thứ tự	Thứ tự biến thể	Vị trí biến thể	Tên biến thể	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)
1	4	16632844	SNP	gcG/gcT	A/A	TT	GG
2	10	16721299	SNP	ccC/ccG	P/P	GG	CC
3	11	16723686	SNP	acG/acA	T/T	GG	AA
4	17	16850100	SNP	ctG/ctA	L/L	AA	GG
5	19	16911171	SNP	gaG/gaA	E/E	AA	GG
6	20	16918401	SNP	gcT/gcA	A/A	AA	TT
7	30	17001369	SNP	tcG/tcA	S/S	AA	GG
8	34	17033896	SNP	ccA/ccT	P/P	TT	AA
9	40	17047096	SNP	gcA/gcT	A/A	AA	TT
10	41	17047369	SNP	ctC/ctT	L/L	AA	GG
11	43	17088291	SNP	ggC/ggT	G/G	TT	CC
12	45	17157660	SNP	Ctg/Ttg	L/L	AA	GG
13	50	17164448	SNP	tcG/tcA	S/S	TT	CC
14	52	17175350	SNP	gcG/gcA	A/A	AA	GG
15	54	17175530	SNP	cgC/cgT	R/R	TT	CC
16	55	17175599	SNP	gtG/gtA	V/V	GG	AA
17	57	17175653	SNP	ctG/ctA	L/L	AA	GG
18	62	17204521	SNP	ccG/ccA	P/P	TT	CC
19	64	17217557	SNP	ccT/ccC	P/P	TT	CC
20	65	17217560	SNP	gaT/gaC	D/D	TT	CC
21	67	17218205	SNP	gcG/gcA	A/A	GG	AA

Số thứ tự	Thứ tự biến thể	Vị trí biến thể	Tên biến thể	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)
22	69	17218262	SNP	gcG/gcA	A/A	GG	AA
23	70	17218271	SNP	gcG/gcA	A/A	GG	AA
24	71	17218274	SNP	ctA/ctG	L/L	AA	GG
25	76	17224392	SNP	tgC/tgT	C/C	TT	CC
26	81	17274324	SNP	cgC/cgT	R/R	AA	GG
27	82	17274372	SNP	gcC/gcT	A/A	AA	GG
28	87	17318299	SNP	gcC/gcT	A/A	TT	CC
29	88	17328082	SNP	gaG/gaA	E/E	TT	CC
30	91	17339840	SNP	ccG/ccA	P/P	AA	GG
31	93	17356069	SNP	acC/acT	T/T	TT	CC
32	97	17356924	SNP	tcC/tcT	S/S	TT	CC
33	101	17358654	SNP	agG/agA	R/R	AA	GG
34	103	17358940	SNP	caG/caA	Q/Q	GG	AA
35	106	17359387	SNP	tcG/tcC	S/S	GG	CC
36	108	17359600	SNP	gtA/gtT	V/V	AA	TT
37	113	17361919	SNP	cgG/cgT	R/R	TT	GG
38	114	17362123	SNP	gaG/gaA	E/E	GG	AA

Bảng 3.6 cho thấy có tổng số 38 biến thể đồng nghĩa đều là biến thể SNP nằm trong vùng exon. Biến thể này làm thay đổi bộ ba mã hóa từ đó làm thay đổi các axit amin.

Tương tự biến thể đồng nghĩa, 62 biến thể sai nghĩa cũng là các biến thể SNP nằm trong vùng exon làm thay đổi bộ ba mã hóa và thay đổi axit amin. Trình tự nucleotit giữa hai haplotype được thể hiện qua bảng 3.7.

Bảng 3.7. Chú giải biến thể sai nghĩa giữa hai haplotype

Số thứ tự	Thứ tự biến thể	Vị trí biến thể	Tên biến thể	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)
1	1	16615477	SNP	atG/atA	M/I	GG	AA
2	2	16615500	SNP	cGt/cTt	R/L	GG	TT
3	3	16615616	SNP	Agg/Tgg	R/W	AA	TT
4	6	16692827	SNP	aCg/aTg	T/M	GG	AA
5	7	16692830	SNP	aCt/aTt	T/I	AA	GG
6	9	16718969	SNP	Gtc/Atc	V/I	TT	CC
7	13	16748183	SNP	cCg/cTg	P/L	GG	AA
8	15	16811690	SNP	aCt/aTt	T/I	AA	GG
9	16	16811708	SNP	gCg/gTg	A/V	GG	AA
10	21	16919033	SNP	aAg/aGg	K/R	AA	GG
11	22	16959756	SNP	gGc/gTc	G/V	GG	TT
12	23	16960905	SNP	caA/caT	Q/H	AA	TT
13	24	16961167	SNP	Agt/Ggt	S/G	AA	GG
14	25	16962253	SNP	Gaa/Aaa	E/K	GG	AA
15	26	16966678	SNP	Gat/Aat	D/N	GG	AA
16	27	16982379	SNP	gCa/gTa	A/V	GG	AA
17	31	17001416	SNP	cCt/cGt	P/R	GG	CC
18	32	17001428	SNP	cGg/cAg	R/Q	AA	GG
19	37	17034312	SNP	Atg/Gtg	M/V	AA	GG
20	38	17034319	SNP	gGg/gAg	G/E	AA	GG
21	42	17085980	SNP	gCc/gTc	A/V	TT	CC

Số thứ tự	Thứ tự biến thể	Vị trí biến thể	Tên biến thể	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)
22	44	17125346	SNP	aAg/aCg	K/T	AA	CC
23	46	17163817	SNP	cGc/cAc	R/H	TT	CC
24	47	17163824	SNP	Cgg/Tgg	R/W	GG	AA
25	49	17163926	SNP	Gaa/Aaa	E/K	TT	CC
26	51	17164506	SNP	gTc/gAc	V/D	AA	TT
27	53	17175375	SNP	Gct/Tct	A/S	GG	TT
28	56	17175601	SNP	cTt/cCt	L/P	TT	CC
29	58	17175777	SNP	Gtg/Atg	V/M	AA	GG
30	59	17175778	SNP	gTg/gCg	V/A	TT	CC
31	60	17203060	SNP	Gtt/Ttt	V/F	AA	CC
32	61	17204382	SNP	gaC/gaA	D/E	TT	GG
33	66	17218199	SNP	atG/atT	M/I	GG	TT
34	68	17218221	SNP	Gcc/Acc	A/T	AA	GG
35	72	17218314	SNP	Gtc/Atc	V/I	AA	GG
36	73	17218414	SNP	cGa/cAa	R/Q	GG	AA
37	74	17220380	SNP	gGt/gAt	G/D	GG	AA
38	75	17220616	SNP	Gcg/Ccg	A/P	GG	CC
39	77	17258929	SNP	ttG/ttT	L/F	AA	CC
40	78	17258984	SNP	cTa/cCa	L/P	GG	AA
41	79	17260450	SNP	tCg/tTg	S/L	GG	AA
42	80	17260559	SNP	Gct/Act	A/T	TT	CC
43	83	17278694	SNP	Ttc/Ctc	F/L	GG	AA

Số thứ tự	Thứ tự biến thể	Vị trí biến thể	Tên biến thể	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)
44	84	17278706	SNP	Ctc/Ttc	L/F	GG	AA
45	85	17290097	SNP	cGc/cAc	R/H	TT	CC
46	86	17313930	SNP	gGg/gAg	G/E	TT	CC
47	89	17328926	SNP	Gat/Aat	D/N	TT	CC
48	90	17334532	SNP	aCt/aAt	T/N	TT	GG
49	94	17356565	SNP	Gag/Aag	E/K	AA	GG
50	96	17356731	SNP	gCc/gTc	A/V	TT	CC
51	98	17357120	SNP	Agg/Ggg	R/G	AA	GG
52	99	17357193	SNP	gAt/gGt	D/G	AA	GG
53	100	17357358	SNP	Cgg/Tgg	R/W	TT	CC
54	102	17358682	SNP	Gtg/Atg	V/M	GG	AA
55	104	17359227	SNP	gCc/gAc	A/D	AA	CC
56	107	17359560	SNP	gAc/gGc	D/G	AA	GG
57	109	17359728	SNP	aCa/aAa	T/K	AA	CC
58	110	17359829	SNP	Act/Gct	T/A	AA	GG
59	111	17360720	SNP	cGt/cAt	R/H	GG	AA
60	115	17362196	SNP	Gag/Aag	E/K	AA	GG
61	116	17362715	SNP	Tgc/Agc	C/S	TT	AA
62	117	17362835	SNP	Ctc/Ttc	L/F	TT	CC

Đối với các biến thể hiệu ứng đột biến khung đọc (bảng 3.8), tám biến thể INDEL nằm trong vùng exon đều không làm thay đổi bộ ba mã hóa cũng như thay đổi axit amin (bảng 3.8).

Bảng 3.8. Chú giải biến thể đột biến khung đọc giữa hai haplotype

Số thứ tự	Thứ tự biến thể	Vị trí biến thể	Tên biến thể	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)
1	5	16639817	INDEL			CGATATTTAGGT GGCACTTTCCCA GTCGATATTTAG GTGGCACTTTCC CAGT	CC
2	8	16692832	INDEL			CTCT	CC
3	28	7000844	INDEL			AGGAGTCGGGG AAGAGGCCACG GACGCCGACAGT TAGGGGATGGA GGCGGCGGCCGC CGGATCCACGTC TCTGAACCTCGT GGAGGCCAGATC CGGCGGTCGCCG GTTGGAGGAGTC GGGAAGAGGC CACGGACGCCGA CAGTTAGGGGAT GGAGGCGGCCG CCGCCGGATCCA	AA

Số thứ tự	Thứ tự biến thể	Vị trí biến thể	Tên biến thể	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)
						CGTCTCTGAACC TCGTGGAGGCCA GATCCGGCGGTC GCCGGTTGG	
4	33	17033835	INDEL			TT	TGTG
5	35	17033997	INDEL			TT	TATA
6	39	17034343	INDEL			TT	TATA
7	105	17359271	INDEL			CTCT	CC
8	112	17361912	INDEL			GG	GTGT

Nhóm biến thể mất bộ ba mã hóa có hai biến thể INDEL nằm tại vị trí 16734616 và 17034001 dự đoán làm thay đổi bộ ba mã hóa và thay đổi axit amin (bảng 3.9).

Bảng 3.9. Chú giải biến thể mất bộ ba mã hóa

Số thứ tự	Thứ tự biến thể	Vị trí biến thể	Tên biến thể	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)
1	12	16734616	INDEL	gcc/-	A5-	AGGC AGGC	AA
2	36	17034001	INDEL	tcatctaag/ -	K/7	TT	TTCATC TAAGTT CATCTA AG

Hai nhóm biến thể thêm bộ ba mã mở đầu (6 biến thể) và thêm bộ ba mã kết thúc (01 biến thể) đều là các biến thể SNP. Trong đó 6 biến thể thêm bộ ba mã mở đầu làm ảnh hưởng cấu trúc protein và biến thể thêm bộ ba mã kết thúc không làm thay đổi cấu trúc protein. Chú giải biến thể được thể hiện qua bảng 3.10.

Bảng 3.10. Chú giải biến thể thêm bộ ba mã mở đầu và thêm bộ ba mã kết thúc

Số thứ tự	Thứ tự biến thể	Vị trí biến thể	Tên biến thể	Hiệu ứng	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)
1	14	16762657	SNP	Thêm bộ ba mã kết thúc	Cag/Tag	Q/8	TT	CC
2	18	16850108	SNP	Thêm bộ ba mã kết thúc	tCa/tGa	S/3	GG	CC
3	29	17001365	SNP	Thêm bộ ba mã kết thúc	tGg/tAg	W/1	GG	AA
4	48	17163902	SNP	Thêm bộ ba mã kết thúc	Caa/Taa	Q/2	AA	GG
5	63	17204583	SNP	Thêm bộ ba mã mở đầu			GG	AA
6	92	17356052	SNP	Thêm bộ ba mã kết thúc	Gga/Tga	G/1	TT	GG
7	95	17356570	SNP	Thêm bộ ba mã kết thúc	tgG/tgA	W/3	GG	AA

Từ kết quả phân tích, chú giải các biến thể, phần mềm CAPSDETECTOR được sử dụng để sàng lọc các SNPs nằm trong vị trí cắt của enzyme giới hạn để xây dựng chi thị phân tử CAPS, phục vụ việc đánh giá kiểu gen của quần thể F₂.

Kết quả sàng lọc 827 biến thể SNP thu được 12 SNP nằm trong vị trí cắt của năm enzyme giới hạn (*SacI*, *DraI*, *EcoRV*, *BamHI*, *Sall*), kết quả thu được thể hiện qua bảng 3.11.

Bảng 3.11. Danh sách SNP nằm trong vị trí cắt của các enzyme giới hạn

Số thứ tự	Vị trí SNP	Tên Locus	Đặc điểm vị trí	Thay đổi bộ ba mã hóa	Thay đổi axit amin	H1 (G6)	H2 (G189)	Enzyme
1	16748183	LOC_Os02g28334	exon	cCg/cTg	P/L	GG	AA	<i>SacI</i>
2	16805351	LOC_Os02g28410	Promoter			GG	TT	<i>DraI</i>
3	16882053	LOC_Os02g28530	Intron			AA	TT	<i>DraI</i>
4	16919772	intergenic				TT	TATA	<i>DraI</i>
5	16961918	LOC_Os02g28670	Promoter			AA	TT	<i>EcoRV</i>
6	17035526	LOC_Os02g28810	Promoter			AA	GG	<i>BamHI</i>
7	17039753	LOC_Os02g28820	Promoter			TT	CC	<i>EcoRV</i>
8	17124682	LOC_Os02g28910	Promoter			TATA	TAT	<i>DraI</i>
9	17161487	LOC_Os02g28980	Promoter			TATA	TT	<i>DraI</i>
10	17205540	LOC_Os02g29040	Promoter			AA	TT	<i>Sall</i>
11	17254246	LOC_Os02g29130	intron			TT	AA	<i>DraI</i>
12	17271258	LOC_Os02g29150	3'-UTR			GG	CC	<i>BamHI</i>

Ghi chú: H1: Haplotype 1, H2: Haplotype 2

Bảng 3.11 cho thấy, 12 SNP nằm trong vị trí cắt của năm enzyme giới hạn chủ yếu nằm vị trí trong vùng promoter (có 7 SNP), chỉ có hai SNP nằm trong vùng intron, một SNP nằm trong vùng 3'-UTR và một SNP nằm trong vùng exon làm thay đổi bộ ba mã hóa từ CCG thành CTG từ đó làm thay đổi axit amin từ Proline (P) sang Leucine (L). Đối với vị trí cắt của enzyme giới hạn

thì có sáu SNP nằm trong vị trí cắt của enzyme giới hạn *DraI*, hai SNP nằm trong vị trí cắt của enzyme *BamHI*, hai SNP nằm trong vị trí cắt của enzyme *EcoRV*, một SNP nằm trong vị trí cắt của enzyme *SacI* và một SNP nằm trong vị trí cắt của enzyme *SalI*.

3.2.2. Phát triển chỉ thị phân tử CAPS

Chỉ thị phân tử CAPS được xây dựng dựa trên sự đa hình đơn nucleotit (SNP) nằm trong vị trí cắt của enzyme giới hạn. Ở những cá thể xảy ra đột biến trong vị trí cắt của enzyme giới hạn (thường là SNPs) thì enzyme sẽ không nhận ra để cắt. Từ đó, mỗi được thiết kế để nhận đoạn DNA chứa vị trí cắt của enzyme giới hạn. Sản phẩm PCR sau đó được cắt bởi enzyme giới hạn sẽ cho các đoạn DNA có chiều dài khác nhau giữa giống có đột biến trong vị trí cắt của enzyme giới hạn (chỉ có một đoạn DNA dài vì enzyme không cắt sản phẩm PCR) và giống không có đột biến (sản phẩm PCR được cắt thành hai mảnh DNA ngắn hơn).

Chỉ thị phân tử CAPS đã được minh chứng có hiệu quả cao trong nhiều nghiên cứu trên nhiều đối tượng. Chẳng hạn như sử dụng chỉ thị phân tử CAPS trong việc xác định đúng loại giống cò dầu mong muốn phục vụ công tác nhân giống, giúp tiết kiệm thời gian, công sức và tiền bạc như công trình của Babu và cộng sự (2017) [14], hay trong việc xác định cá thể F_1 trong các quần thể lai loài dâu tằm của Arora và cộng sự (2017) [13]. Vì vậy, để phát triển chỉ thị phân tử CAPS phục vụ công tác chọn lọc các dòng F_2 đồng hợp tử mang QTL9 của bố và các dòng F_2 đồng hợp tử mang QTL9 của mẹ, công nghệ chụp gen đã được ứng dụng để xác định các SNP nằm trên vùng QTL9 của hai giống bố mẹ, từ đó sàng lọc ra các SNPs nằm trong vùng cắt của enzyme giới hạn.

Dựa vào 12 SNPs tìm thấy trong vùng cắt của các enzyme giới hạn (Bảng 3.11), các môi đã được thiết kế xung quanh vị trí của SNPs nhằm

khuếch đại vùng chứa SNPs, sao cho khi cắt sản phẩm PCR sẽ tạo ra đa hình về chiều dài đoạn cắt.

Các môi thiết kế dựa trên trình tự của giống Nipponbarre trên trang MSU Rice Genome Annotation (Osa1) Release 7 (<http://rice.plantbiology.msu.edu>) [136], sử dụng phần mềm Primer3 plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) [137].

Kết quả thu được 12 chỉ thị phân tử CAPS thể hiện qua bảng 3.12.

Bảng 3.12. Danh sách 12 chỉ thị phân tử CAPS và trình tự môi

Tên	Vị trí	Tên Locus	Đặc điểm vị trí	Trình tự môi (5'-3')	Kích thước sản phẩm PCR
CAPS 1	16747814	LOC_Os02g28334	UTR_3'	F- CTCCAGCATCTCACCCCTCTC R- CTCCACCATGAATAACCTGCA	357
CAPS 2	16805081	LOC_Os02g28410	Promoter	F- GTTACTCCATCCGTCCCAA R- GAATGCTCTCGTCGGATGAT	877
CAPS 3	16882053	LOC_Os02g28530	intron	F- TGGTTTGGGTTACTCGGAAG R- TTACCTGGAGTGCAAGGTCC	689
CAPS 4			intragenic	F- TTGTTTCGTGATTCGCACGG R- CGTGCGAACCACGAAG	852
CAPS 5	16961918	LOC_Os02g28670	Promoter	F- GGCAGCATCAGCAAGCATA R- GGATTAGCTCCCTACCACGC	1037
CAPS 6	17034718	LOC_Os02g28810	Promoter	F- CTGAATCCGTCCAGCAAGGT R- TGCGAGAGGAACGAAACGAA	577
CAPS 7	17039112	LOC_Os02g28820	Promoter	F- TGGTGATTACGGCGTTTGGT R- TTCGGCAATACGTCACTA	1358
CAPS 8	17124657	LOC_Os02g28910	Promoter	F- CACGCAGTTTAGCATGACGG R- TTCCCGTACTCTCGTCACCT	1052
CAPS 9	17156453	LOC_Os02g28980	UTR_3'	F- TGAGAAGCGAAACGCACCTT R- GCTTCCTCTGCTGTCGGTAA	1710

Tên	Vị trí	Tên Locus	Đặc điểm vị trí	Trình tự môi (5'-3')	Kích thước sản phẩm PCR
CAPS 10	17196458	LOC_Os02g29040	intron	F- GTCTGCGCCCACATTTTAGT R- GCCTTTCTCTCACCTTGCAC	1909
CAPS 11	17246451	LOC_Os02g29130	intron	F- TGGATTGCATGGTTTAGCTG R- ATGCATTGAGAGGGACCTTG	898
CAPS 12	17271233	LOC_Os02g29150	UTR_3'	F- TGGTGGATTAATGCTGTGGA R- GCCACATTTAGATGTGTAA	1103

Mười hai chỉ thị phân tử CAPS được xây dựng từ 12 SNPs nằm trong vị trí cắt của 5 enzyme giới hạn. Trình tự vị trí cắt và kích thước sản phẩm cắt của các enzyme giới hạn thể hiện qua bảng 3.13 sau:

Bảng 3.13. Trình tự vị trí cắt và kích thước các sản phẩm cắt của 5 enzyme giới hạn sử dụng để cắt chỉ thị phân tử CAPS

Chỉ thị phân tử CAPS	SNP (H1/H2)	Enzyme	Trình tự vị trí cắt của enzyme giới hạn		Kích thước sản phẩm cắt bằng enzyme giới hạn	
			<i>G6 (H1)</i>	<i>G189 (H2)</i>	<i>G6 (H1)</i>	<i>G189 (H2)</i>
CAPS 1	G/A	<i>SacI</i>	GAGCTC	<u>A</u> GCTC	134 & 223	357
CAPS 2	T/G	<i>DraI</i>	TTTAAA	TT <u>G</u> AAA	346 & 531	877
CAPS 3	A/T	<i>DraI</i>	TT <u>A</u> AAA	TTTAAA	689	473 & 216
CAPS 4	T/TA	<i>DraI</i>	TTTAAA	TT <u>A</u> AAA	601 & 251	852
CAPS 5	A/T	<i>EcoRV</i>	GATATC	G <u>T</u> TATC	643 & 394	1037
CAPS 6	G/A	<i>BamHI</i>	G <u>A</u> ATCC	GGATCC	577	304 & 273
CAPS 7	T/C	<i>EcoRV</i>	GATATC	GAT <u>A</u> CC	1321 & 127	1358
CAPS 8	TA/T	<i>DraI</i>	TTTAAA	TTT <u>T</u> AA	273 & 779	1052
CAPS 9	TA/T	<i>DraI</i>	TTTAAA	TTTT <u>T</u> A	1027 & 683	1710

Chỉ thị phân tử CAPS	SNP (H1/H2)	Enzyme	Trình tự vị trí cắt của enzyme giới hạn		Kích thước sản phẩm cắt bằng enzyme giới hạn	
			<i>G6 (H1)</i>	<i>G189 (H2)</i>	<i>G6 (H1)</i>	<i>G189 (H2)</i>
CAPS 10	A/T	<i>SalI</i>	GTCGAC	GTCGTC	886&1023	1909
CAPS 11	T/A	<i>DraI</i>	TTTAAA TTTAAA	TTTAAT TTTAAA	226&333&339	559&339
CAPS 12	G/C	<i>BamHI</i>	GGATGC	GGATCC	1103	1041&62

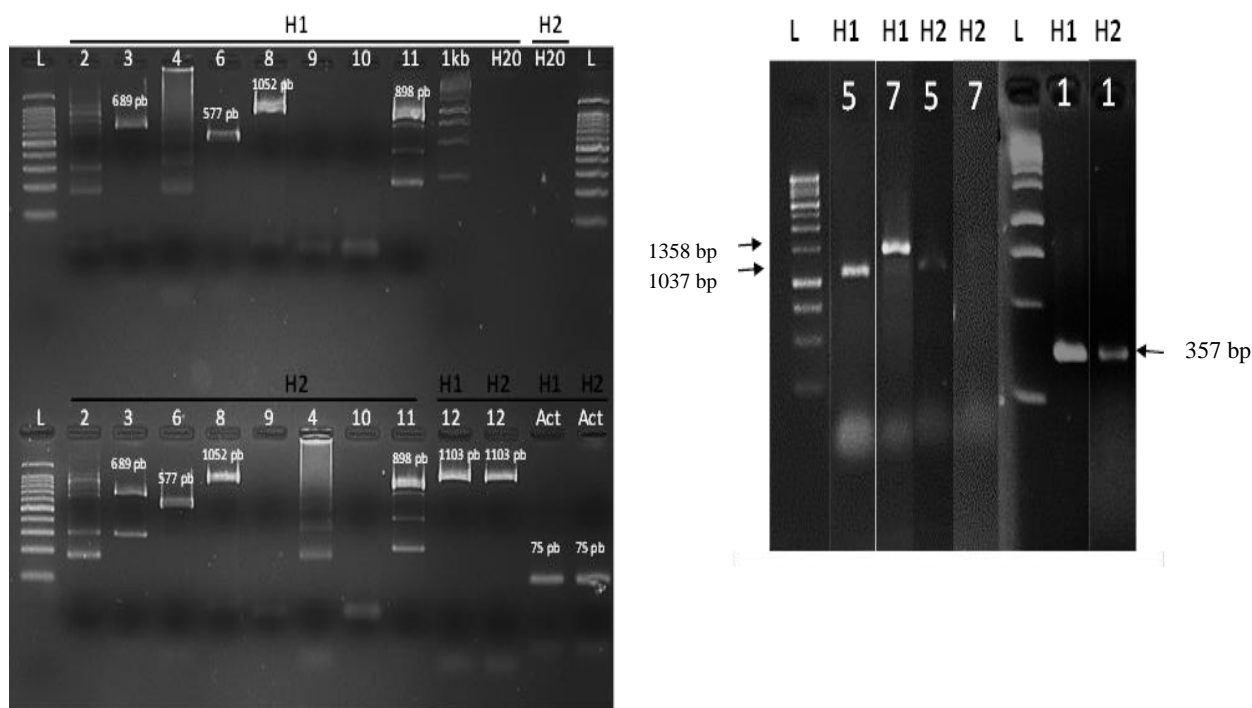
Chữ gạch chân: điểm đột biến trong vị trí cắt của enzyme giới hạn.

Các chỉ thị phân tử CAPS và 5 enzyme giới hạn này sẽ được sử dụng để xác định các dòng F₂ đồng hợp tử mang QTL9 của bố và các dòng F₂ đồng hợp tử mang QTL9 của mẹ phục vụ đánh giá kiểu hình cấu trúc bông lúa trong quần thể F₃ ngoài đồng ruộng.

3.3. Chọn lọc các cây F₂ mang QTL9 đồng hợp thuộc hai haplotype bố hoặc mẹ bằng chỉ thị phân tử CAPS

3.3.1. Xác định chỉ thị phân tử CAPS cho đa hình giữa hai giống lúa bố mẹ

Trước khi sử dụng mỗi CAPS để xác định sự phân ly trong quần thể F₂, các môi cần được kiểm tra sự đa hình ở hai giống bố (G189) (H1) và mẹ (G6) (H2) (hai haplotype) bằng phản ứng PCR và cắt enzyme giới hạn. Kết quả điện di phản ứng PCR để kiểm tra độ đặc hiệu với 12 môi CAPS ở hai giống bố mẹ (H1 và H2) được thể hiện trong hình 3.10.



Hình 3.10. Kết quả phản ứng PCR thử độ đặc hiệu của các môi CAPS ở hai giống G6 (H1) và G189 (H2)

(Các giếng số từ 1-12 tương ứng với 12 chỉ thị phân tử CAPS; L: Thang chuẩn DNA Low Gene Ruler 100 bp)

Hình 3.10 cho thấy, sáu cặp môi CAPS cho kết quả đặc hiệu chỉ một băng DNA duy nhất, đúng với kích thước mong đợi ở cả hai giống bố mẹ: CAPS 1 (357 bp), CAPS 5 (1037 bp), CAPS 6 (577 bp), CAPS 7 (1358 bp), CAPS 8 (1052 bp), CAPS 12 (1103 bp).

Kết quả PCR với môi CAPS 2 không đặc hiệu, xuất hiện nhiều băng DNA với kích thước khác nhau nhưng không có băng nào đúng với kích thước mong đợi (877 bp), hình ảnh điện di giống nhau ở cả hai haplotype.

Với môi CAPS 3 chỉ có một băng đặc hiệu đúng kích thước mong đợi (689 bp) ở H1, tuy nhiên ở H2, ngoài băng đúng kích thước mong đợi còn xuất hiện một băng không đặc hiệu, có kích thước nhỏ hơn.

Kết quả PCR với môi CAPS 4 cho hai băng có kích thước khoảng 200 và 300 bp, không đúng với kích thước mong đợi (852 bp), ở cả hai haplotype.

Phản ứng PCR với mỗi CAPS 9 và CAPS 10 không thu được băng nào ở cả hai haplotype.

Kết quả PCR với mỗi CAPS 11 cho nhiều băng DNA với kích thước khác nhau nhưng, trong đó có băng đúng với kích thước mong đợi (898 bp), hình ảnh điện di giống nhau ở cả hai haplotype.

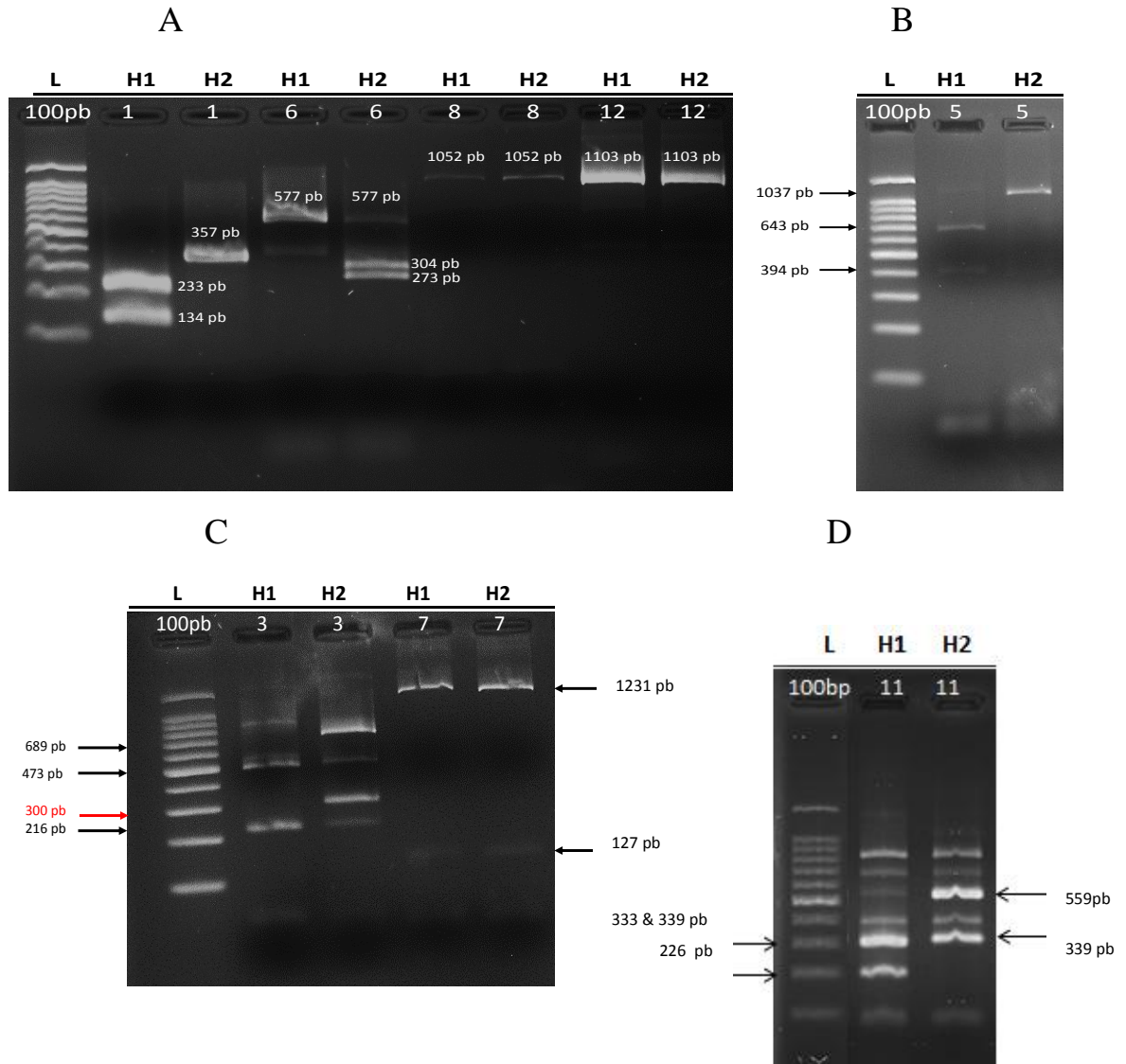
Như vậy, qua kết quả PCR kiểm tra độ đặc hiệu của 12 mỗi CAPS thu được 8 chỉ thị CAPS có băng đúng với kích thước mong đợi ở cả hai giống bố mẹ H1 (G6) và H2 (G189) bao gồm: CAPS 1, CAPS 3, CAPS 5, CAPS 6, CAPS 7, CAPS 8, CAPS 11, CAPS 12.

Sản phẩm PCR của tám mỗi CAPS trên tiếp tục được xử lý bằng các enzyme giới hạn tương ứng: CAPS 1 cắt bằng enzyme *SacI*; CAPS 3, CAPS 8, CAPS 11 cắt bằng enzyme *DraI*, CAPS 5 cắt bằng enzyme *EcoRV*; CAPS 6, CAPS 7, CAPS 11 cắt bằng enzyme giới hạn *BamHI*. Kết quả thể hiện qua hình 3.11.

Hình 3.11 cho thấy: Sản phẩm PCR với mỗi CAPS 1 được cắt bằng enzyme giới hạn *SacI*. Kết quả điện di cho sự đa hình giữa hai haplotype. Sản phẩm PCR của H1 bị cắt thành hai mảnh đúng với kích thước mong đợi (134 bp và 233 bp), trong khi ở H2, SNPs xảy ra trong vị trí cắt của *SacI* nên sản phẩm PCR (357 bp) không bị cắt (hình 3.11A).

Sản phẩm PCR với mỗi CAPS 3 được cắt bằng enzyme giới hạn *DraI*. Vị trí cắt của enzyme ở H1 dự đoán không có SNP nên sản phẩm PCR của H1 bị cắt hoàn toàn thành 2 băng DNA đúng kích thước mong đợi (216 bp và 473 bp) (hình 3.11C). Mặc dù dự đoán SNP xảy ra trong vị trí cắt của *DraI* ở H2 nhưng sản phẩm cắt enzyme giới hạn của H2 cũng có hai băng DNA đúng kích thước mong đợi. Tuy nhiên sản phẩm PCR của H2 không bị cắt hoàn toàn, có thể enzyme *DraI* cắt không tốt, độ đặc hiệu không cao. Ngoài ra ở H2 còn có một băng DNA khoảng 300 bp là sản phẩm không đặc hiệu của phản

ứng PCR, do đó có thể phân biệt H1 và H2 nhờ sự khác biệt của băng DNA không đặc hiệu này (Hình 3.11C).



Hình 3.11. Kết quả điện di sản phẩm cắt bằng enzyme giới hạn.

A, Kết quả cắt enzyme giới hạn của sản phẩm PCR với môi CAPS 1, 6, 8 và 12. B, Kết quả cắt enzyme giới hạn của sản phẩm PCR với môi CAPS 5. C, Kết quả cắt enzyme giới hạn của sản phẩm PCR với môi CAPS 3, CAPS 7. D, Kết quả cắt enzyme giới hạn của sản phẩm PCR với môi CAPS 11. (L: Thang chuẩn DNA Low Gene Ruler 100 bp)

Sản phẩm PCR với môi CAPS 5 được cắt bằng enzyme giới hạn *EcoRV*. Kết quả điện di cho sự đa hình giữa hai haplotype. Sản phẩm PCR của H1 bị cắt thành hai mảnh đúng với kích thước mong đợi (394 bp và 643 bp), trong khi ở H2, SNPs xảy ra trong vị trí cắt của *EcoRV* nên sản phẩm PCR (1037 bp) không bị cắt (Hình 3.11B).

Sản phẩm PCR với môi CAPS 6 được cắt bằng enzyme giới hạn *BamHI*. Kết quả điện di cho sự đa hình giữa hai haplotype. Sản phẩm PCR của H2 bị cắt thành 2 mảnh đúng với kích thước mong đợi (304 bp và 273 bp), trong khi ở H1, SNPs xảy ra trong vị trí cắt của *BamHI* nên sản phẩm PCR (577 bp) không bị cắt (Hình 3.11A).

Sản phẩm PCR với môi CAPS 7 được cắt bằng enzyme giới hạn *EcoRV*. Sản phẩm PCR của H1 bị cắt thành hai mảnh đúng với kích thước mong đợi (1231 bp và 127 bp). Tuy nhiên, trái với dự đoán, sản phẩm PCR của H2 cũng bị cắt thành hai mảnh giống như ở H1, có thể việc dự đoán SNP xuất hiện trong vị trí cắt của enzyme ở H2 không chính xác (Hình 3.11C). Trường hợp ngược lại xảy ra với CAPS 8 và CAPS 12. Sản phẩm PCR với môi CAPS 8 được cắt bằng enzyme giới hạn *DraI* còn sản phẩm PCR với môi CAPS 12 được cắt bằng enzyme giới hạn *BamHI*. Tuy nhiên phản ứng cắt đã không diễn ra như dự đoán ở H1 mà chỉ có 1 băng DNA đúng với kích thước sản phẩm PCR (1052 bp đối với CAPS 8 và 1103bp đối với CAPS 12). Sản phẩm PCR ở H2 không bị cắt đúng như dự đoán. Có thể việc dự đoán vị trí cắt của enzyme *DraI* và *BamHI* trong vùng CAPS 8 và CAPS 12 ở H1 không bị đột biến là không chính xác.

Sản phẩm PCR với môi CAPS 11 được cắt bằng enzyme giới hạn *DraI*. Có 2 vị trí cắt của enzyme trong sản phẩm PCR của H1 nên sản phẩm PCR được cắt thành 3 mảnh (226, 333 và 339 bp). Trong khi chỉ có 1 vị trí cắt ở H2 nên sản

phẩm PCR của H2 được cắt thành 2 mảnh đúng với kích thước mong đợi (339 bp và 559 bp) (Hình 3.11D).

Như vậy, qua kết quả cắt bằng enzyme giới hạn của tám chỉ thị CAPS cho thấy có bốn chỉ thị CAPS cho sự đa hình giữa hai giống bố mẹ và bao phủ vùng QTL9 đó là CAPS 1, CAPS 5, CAPS 6 và CAPS 11. Các chỉ thị này có thể được sử dụng để xác định các cây F_2 mang QTL9 đồng hợp tử.

3.3.2. Chọn lọc các cây F_2 đồng hợp tử mang QTL9 của bố hoặc mẹ bằng chỉ thị phân tử CAPS

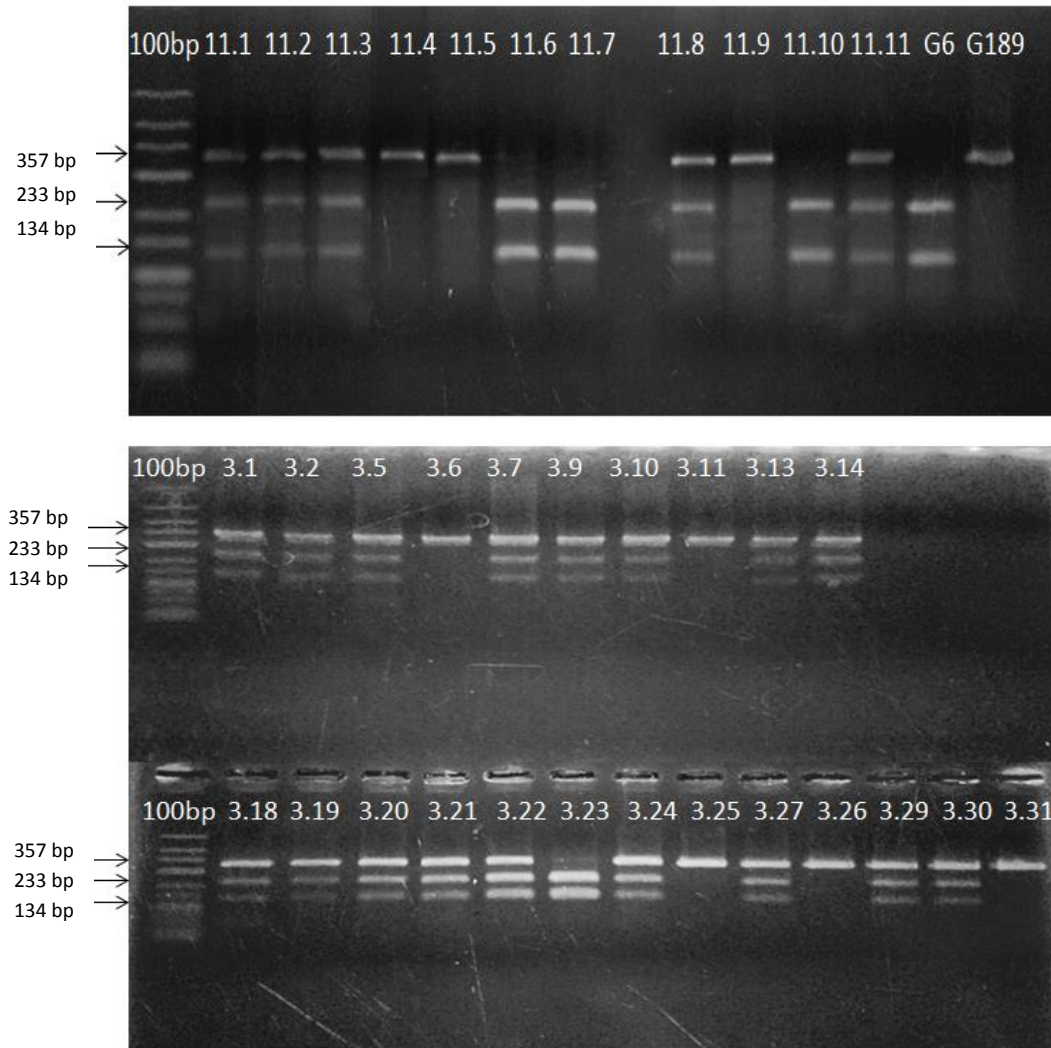
Trong bốn chỉ thị phân tử CAPS cho sự đa hình giữa hai giống bố mẹ và bao phủ vùng QTL9: CAPS 1, CAPS 5, CAPS 6 và CAPS 11, trong đó có CAPS 5 và CAPS 6 có vị trí gần nhau nên đề tài chỉ chọn ba chỉ thị CAPS để chọn lọc các dòng F_2 đồng hợp tử mang QTL9 của giống mẹ (G6) và các dòng F_2 đồng hợp tử mang QTL9 của giống bố (G189): CAPS 1, CAPS 5 và CAPS 11.

275 cây F_2 được chạy phản ứng PCR bằng các môi CAPS, sau đó cắt sản phẩm PCR bằng các enzyme tương ứng: CAPS 1 với enzyme *SacI*, CAPS 5 với enzyme *EcoRV* và CAPS 11 với enzyme *DraI*, sản phẩm điện di trên gel agarose 2%. Các kết quả thu được như sau:

Đối với CAPS 1: Kết quả phản ứng PCR và cắt bằng enzyme *SacI*, sau khi điện di trên gel agarose 2% xác định được các cây mang QTL9 đồng hợp tử của cây mẹ (G6) xuất hiện hai băng có kích thước tương ứng 134 bp và 233 bp. Các cây mang QTL9 đồng hợp tử của cây bố (G189) chỉ thu được một băng DNA duy nhất tương ứng với kích thước 357 bp. Các cây dị hợp tử thu được ba băng có kích thước 134 bp, 233 bp và 357 bp.

Hình 3.12 minh họa một số kết quả thu được với cặp môi CAPS 1. Những cây có hai băng DNA kích thước 134 bp và 233 bp (11.6; 11.7; 11.10; 3.23) là cây mang QTL9 đồng hợp của cây mẹ (G6). Những cây chỉ có một

băng DNA duy nhất ở kích thước 357 bp (11.4; 11.5; 3.6; 3.11; 3.25; 3.26; 3.31) là cây đồng hợp tử mang QTL9 của cây bố (G189). Những cây có cả ba băng DNA kể trên là cây dị hợp tử mang cả QTL9 của bố và mẹ.

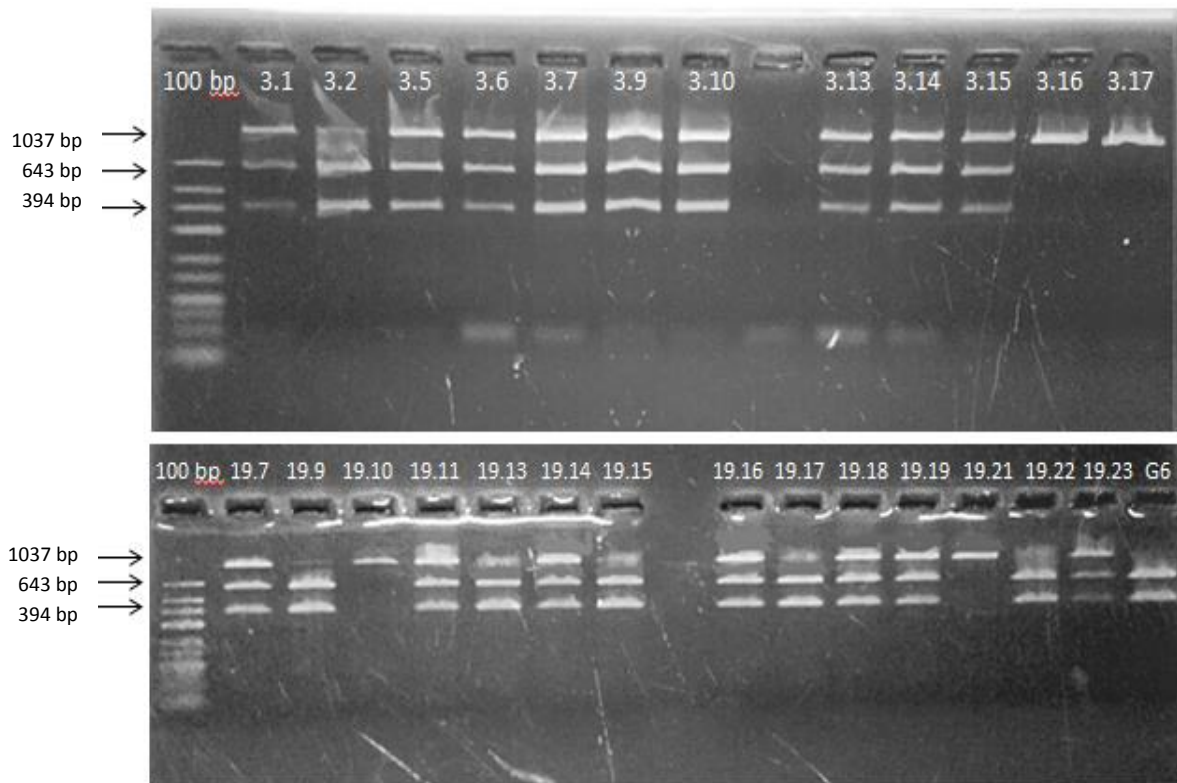


Hình 3.12. Kết quả phân tích một số cây F₂ với chỉ thị CAPS 1

Kết quả phân tích 275 cây F₂ thu được 74 cây F₂ mang QTL9 đồng hợp của cây mẹ (G6), 66 cây mang QTL9 đồng hợp của cây bố (G189) và 135 cây dị hợp tử mang cả hai QTL9 của bố và mẹ (danh sách các cây đồng hợp tử thể hiện qua bảng 3.14).

Đối với CAPS 5: 275 mẫu DNA của quần thể F₂ được phân tích với mỗi CAPS 5 bằng phản ứng PCR, sau đó cắt sản phẩm PCR bằng enzyme

EcoRV thu được các cây đồng hợp tử của cây mẹ (G6) xuất hiện hai băng với kích thước 394 bp và 643 bp. Các cây đồng hợp tử của cây bố (G189) chỉ xuất hiện một băng duy nhất ở kích thước 1037 bp. Các cây dị hợp tử xuất hiện ba băng ở các kích thước 394 bp, 643 bp và 1087 bp (hình 3.13).



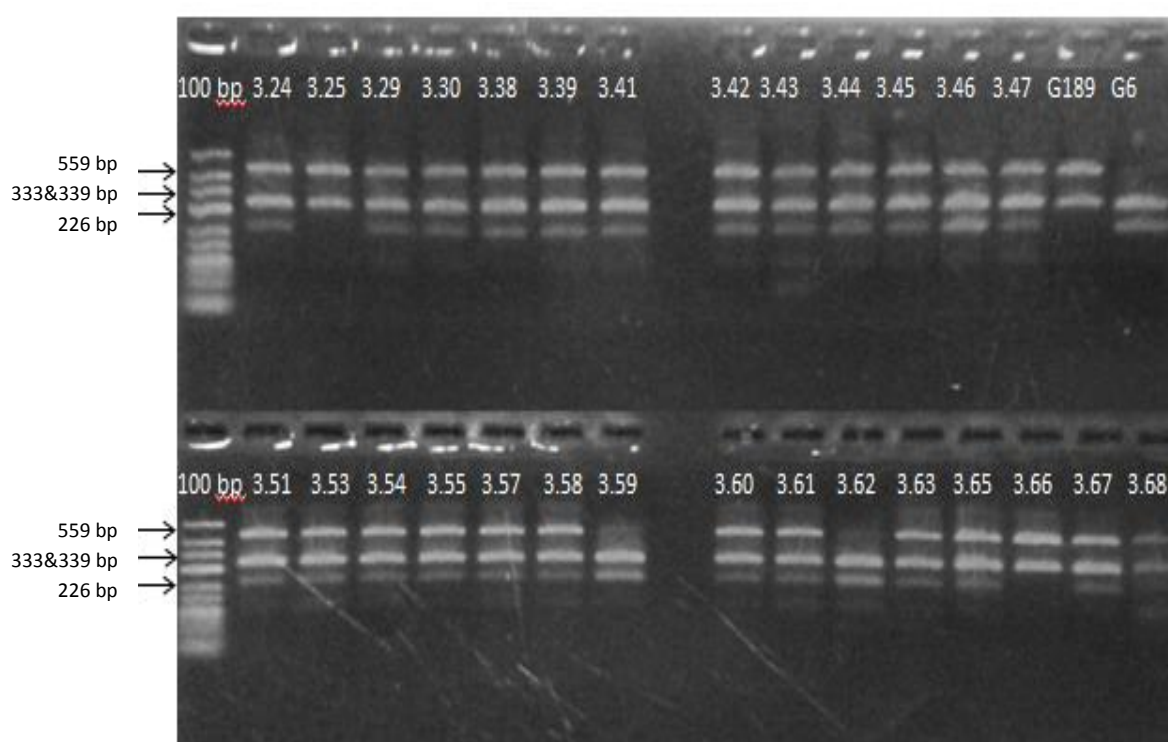
Hình 3.13. Kết quả phân tích một số cây F₂ với chỉ thị CAPS 5

Hình 3.13 là kết quả điện di sản phẩm cắt enzyme của một số cây F₂. Những cây chỉ có hai băng DNA kích thước 394 bp và 643 bp (19.9; 19.13; 19.15; 19.17; 19.22) là cây mang QTL9 đồng hợp của cây mẹ (G6). Những cây chỉ có một băng DNA duy nhất ở kích thước 1037 bp (3.16; 3.17; 19.10; 19.21) là cây mang QTL9 đồng hợp của cây bố (G189). Những cây có cả ba băng DNA kể trên là cây dị hợp mang cả QTL9 của bố và mẹ.

Kết quả phân tích 275 cây F₂ thu được 74 cây F₂ mang QTL9 đồng hợp của cây mẹ (G6), 66 cây mang QTL9 đồng hợp của cây bố (G189) và 135 cây dị hợp tử mang cả hai QTL9 của bố và mẹ. Kết quả này hoàn toàn trùng lặp

với kết quả thu được sử dụng CAPS 1 (danh sách các cây đồng hợp tử thể hiện qua bảng 3.14).

Tương tự 275 mẫu DNA của quần thể F_2 được phân tích kiểu gen với môi **CAPS 11**, cắt bằng enzyme *DraI* thu được các cây F_2 mang QTL9 đồng hợp của cây mẹ (G6) chỉ có ba băng kích thước 226 bp và 333 và 339 bp; các cây chỉ có hai băng DNA với kích thước 339 bp và 559 bp là cây mang QTL9 đồng hợp tử của bố (G189) (hình 3.14).



Hình 3.14. Kết quả phân tích một số cây F_2 với chỉ thị CAPS 11

Hình 3.14 là kết quả điện di của sản phẩm PCR được cắt bằng enzyme *DraI* của một số cây F_2 . Những cây mang QTL9 đồng hợp của cây mẹ (G6) bao gồm 3.46; 3.59; 3.62 có ba băng DNA kích thước 226 bp; 333 và 339 bp. Những cây chỉ có hai băng DNA với kích thước 339 bp và 559 bp (3.25; 3.66) là cây mang QTL9 đồng hợp của cây bố (G189). Những cây có cả bốn băng DNA kể trên là các cây dị hợp mang cả QTL9 của bố và mẹ.

Kết quả phân tích 275 cây F₂ thu được 74 cây F₂ mang QTL9 đồng hợp của cây mẹ (G6), 66 cây mang QTL9 đồng hợp của cây bố (G189) và 135 cây F₂ dị hợp tử mang cả hai QTL9 của bố và mẹ. Kết quả này cũng hoàn toàn trùng lặp với kết quả thu được sử dụng CAPS 1 và CAPS 5 (danh sách các cây đồng hợp tử thể hiện qua bảng 3.14).

Như vậy, nhờ chỉ thị phân tử CAPS được thiết kế bao phủ vùng QTL9 đã giúp xác định sự phân ly kiểu gen của QTL9, cụ thể là xác định các dòng đồng hợp tử mang QTL9 của bố hoặc mẹ ngay trong quần thể F₂ thay vì đánh giá đến thế hệ F₉ và F₁₀ như trong nghiên cứu của Zhao và các cộng sự (2016) [126] trong quần thể RILs.

275 cá thể F₂ được phân tích với 3 chỉ thị phân tử CAPS 1, CAPS 5 và CAPS 11 thu được 74 cây F₂ đồng hợp tử mang kiểu gen QTL9 của mẹ (G6), 66 cây F₂ đồng hợp tử mang kiểu gen QTL9 của bố (G189), chiếm tỷ lệ 50,9%. Danh sách các cây đồng hợp tử được thể hiện qua bảng 3.14 sau:

Bảng 3.14. Danh sách cây F₂ đồng hợp tử được xác định bằng 3 chỉ thị phân tử CAPS 1, CAPS 5, CAPS 11

Các cây F ₂ mang QTL9 đồng hợp tử của cây mẹ G6				Các cây F ₂ mang QTL9 đồng hợp tử của cây bố G189			
STT	Ký hiệu cây	STT	Ký hiệu cây	STT	Ký hiệu cây	STT	Ký hiệu cây
1	1.2	38	2.89	1	1.6	38	2.129
2	1.10	39	2.97	2	1.8	39	2.133
3	1.11	40	2.99	3	1.9	40	2.136
4	1.16	41	2.100	4	1.13	41	3.11
5	1.20	42	2.113	5	1.14	42	3.16
6	1.29	43	2.121	6	1.15	43	3.17
7	1.31	44	2.123	7	1.18	44	3.25

Các cây F₂ mang QTL9 đồng hợp tử của cây mẹ G6				Các cây F₂ mang QTL9 đồng hợp tử của cây bố G189			
8	1.33	45	2.132	8	1.22	45	3.26
9	1.38	46	2.137	9	1.23	46	3.31
10	1.42	47	3.2	10	1.28	47	3.36
11	2.5	48	3.22	11	1.32	48	3.49
12	2.12	49	3.23	12	1.40	49	3.66
13	2.13	50	3.33	13	1.45	50	3.69
14	2.18	51	3.34	14	1.49	51	3.73
15	2.19	52	3.37	15	2.2	52	3.74
16	2.25	53	3.46	16	2.8	53	11.4
17	2.28	54	3.50	17	2.9	54	11.5
18	2.32	55	3.59	18	2.10	55	11.9
19	2.34	56	3.62	19	2.16	56	15.1
20	2.35	57	3.70	20	2.20	57	18.4
21	2.37	58	3.71	21	2.36	58	18.12
22	2.38	59	11.6	22	2.44	59	18.13
23	2.40	60	11.7	23	2.49	60	18.17
24	2.42	61	11.10	24	2.61	61	18.19
25	2.46	62	17.2	25	2.62	62	19.1
26	2.48	63	17.4	26	2.75	63	19.5
27	2.53	64	18.1	27	2.77	64	19.6
28	2.57	65	18.7	28	2.87	65	19.10
29	2.59	66	18.8	29	2.92	66	19.21
30	2.66	67	18.16	30	2.93		
31	2.68	68	18.21	31	2.94		
32	2.71	69	18.24	32	2.104		

Các cây F ₂ mang QTL9 đồng hợp tử của cây mẹ G6				Các cây F ₂ mang QTL9 đồng hợp tử của cây bố G189	
33	2.72	70	19.9	33	2.107
34	2.74	71	19.13	34	2.109
35	2.79	72	19.15	35	2.112
36	2.83	73	19.17	36	2.122
37	2.86	74	19.22	37	2.124
Tổng cộng: 74 cây				Tổng cộng: 66 cây	

Các cây F₂ đồng hợp tử sẽ được phát triển tạo quần thể F₃, phân tích kiểu hình cấu trúc bông của quần thể F₃ so sánh với hai giống bố mẹ để làm sáng tỏ được vai trò của QTL9 liên quan đến cấu trúc bông lúa.

3.4. Phân tích kiểu hình cấu trúc bông của hai quần thể F₃ thuộc hai haplotype bố hoặc mẹ

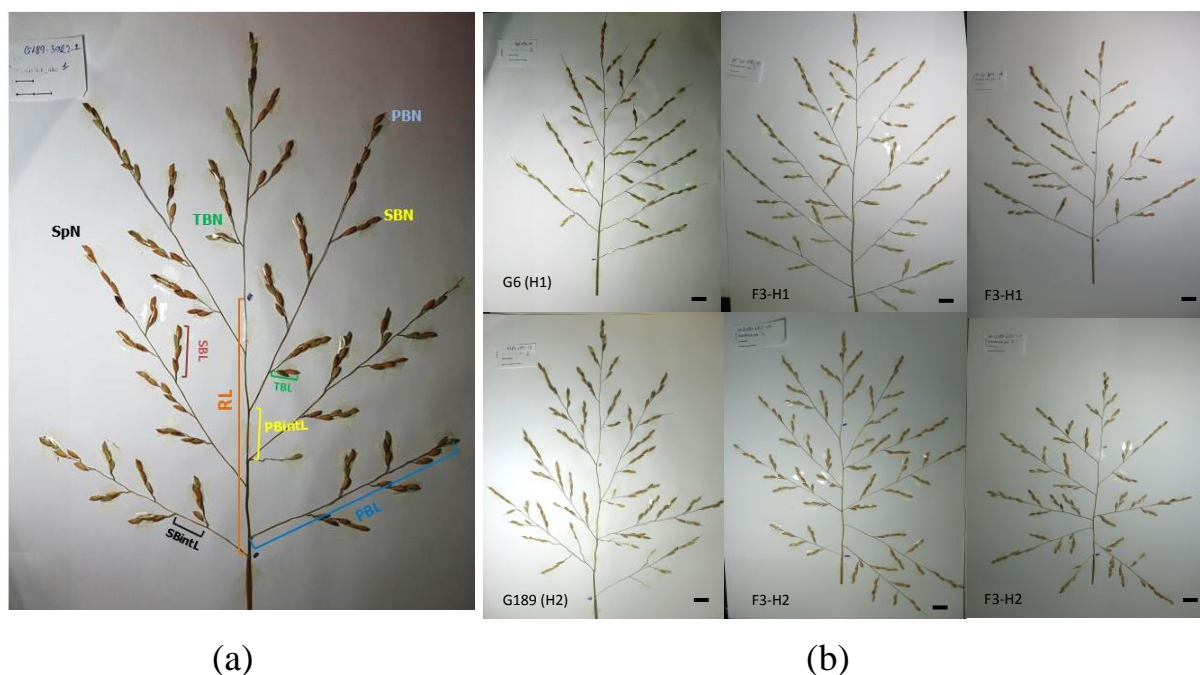
QTL9 được xác định liên kết với hai tính trạng số gié cấp hai/bông và số hạt/bông bằng phương pháp phân tích liên kết trên toàn hệ gen (GWAS) của nhóm nghiên cứu Tạ Kim Nhung và các cộng sự từ năm 2013 - 2014 (công bố năm 2018). Tuy nhiên, kết quả phân tích này dựa trên các thuật toán thống kê. Vì vậy, để kiểm chứng xem QTL9 có thực sự liên kết với hai tính trạng này hay không thông qua các quần thể lai tái tổ hợp, chọn lọc và tạo quần thể lai đồng hợp tử mang QTL9 sau đó đánh giá kiểu hình ngoài đồng ruộng. So sánh kiểu hình cấu trúc bông lúa của quần thể đồng hợp tử với quần thể bố mẹ, từ đó làm sáng tỏ vai trò của QTL9 liên quan đến hai tính trạng số gié cấp hai/bông và số hạt/bông.

Có nhiều quần thể có thể sử dụng để nghiên cứu QTL như quần thể F₂ của Bevan và cộng sự, 2017 [25]; hay quần thể đơn bội DHs của nhóm nghiên cứu Bao và cộng sự, 2002 [21], Hill và cộng sự, 2019 [46]; hay phát triển

quần thể NILs theo Bai Xufeng và cộng sự, 2012 [18]; quần thể lai trở lại cận giao tái tổ hợp BILs của Thomson và cộng sự, 2003 [107] và nhóm nghiên cứu của Peng và cộng sự, 2014 [87]... Tuy nhiên, để phát triển quần thể NILs hay BILs thường tốn nhiều thời gian và công sức. Vì vậy, đề tài đã phát triển quần thể lai RILs và sử dụng chỉ thị phân tử CAPS để chọn lọc dòng F₃ đồng hợp tử mang QTL9 của bố hoặc mẹ thay vì chọn dòng đồng hợp tử ở thế hệ F₇ như phương pháp truyền thống của Marathi và cộng sự, 2012 [76].

Tổng số 74 dòng F₂ đồng hợp tử mang QTL9 của cây mẹ G6 (H1) và 66 dòng F₂ đồng hợp tử mang QTL9 của cây bố G189 (H2) đã được xác định (bảng 3.9). Chọn lọc 49 dòng đồng hợp tử thu được nhiều hạt với mỗi kiểu haplotype H1 và H2 để đánh giá kiểu hình cấu trúc bông của quần thể F₃ ngoài đồng ruộng. Thí nghiệm được thực hiện vào vụ xuân năm 2019, gồm 98 ô thí nghiệm tương ứng với 98 dòng đồng hợp tử và hai ô đối chứng G6, G189 được bố trí ngẫu nhiên, mỗi ô gồm 16 cây được cấy thành 4 hàng, mỗi hàng có 4 cây.

Tại mỗi ô thí nghiệm, 9 bông từ 3 cây ngẫu nhiên ở giữa ô được thu thập (mỗi cây thu 3 bông). Bông lúa sau khi thu, được dán cố định trên khổ giấy A₃, sử dụng phần mềm PTRAP được phát triển bởi nhóm nghiên cứu của AL-Tam và các cộng sự thuộc Viện Nghiên cứu và Phát triển Pháp - IRD [12] để đo đếm các chỉ tiêu quan tâm (Hình 3.15).



Hình 3.15. Phân tích định lượng các tính trạng cấu trúc bông lúa
 (a) Định lượng thành phần cấu trúc bông lúa bằng phần mềm P-Trap các chỉ tiêu RL: chiều dài trục bông; PBN: số gié cấp một/bông; SBN: số gié cấp hai/bông; TBN: số gié cấp ba/bông; SpN: số hạt/bông; PBL: chiều dài gié cấp một; SBL: chiều dài gié cấp hai; TBL: chiều dài gié cấp ba; PBintL: khoảng cách giữa các gié cấp một; SBintL: khoảng cách giữa các gié cấp hai. (b) (hàng trên) hình ảnh bông lúa cây mẹ G6 (H1) và 2 bông của cây F₃ đồng hợp tử mang QTL9 của cây mẹ, (hàng dưới) hình ảnh bông lúa cây bố G189 (H2) và 2 bông lúa của cây F₃ đồng hợp tử mang QTL9 của bố (Ảnh chụp tại Viện Di truyền Nông nghiệp năm 2019).

Thanh tỷ lệ: 2 cm

Tổng số 900 bông lúa đã được phân tích cấu trúc bông lúa bằng phần mềm P-Trap các chỉ tiêu: Chiều dài bông (RL), số gié cấp một (PBN), chiều dài gié cấp một (PL), số gié cấp hai (SBN), chiều dài gié cấp hai (SBL), số gié cấp ba (TBN), khoảng cách giữa các gié cấp một (PBintL), khoảng cách giữa các gié cấp hai (SBintL) và số hạt/bông (SpN). Ngoài ra, các chỉ tiêu chiều cao cây, số nhánh hữu hiệu, thời gian ra hoa cũng được đo đếm trực tiếp trên 150 cây thí nghiệm.

Kết quả đo đếm các chỉ tiêu thể hiện qua bảng 3.15 như sau:

Bảng 3.15. Kết quả đo đếm thành phần cấu trúc bông lúa bằng phần mềm P-Trap và các chỉ tiêu liên quan đến năng suất lúa của quần thể bố mẹ và quần thể F₃ đồng hợp tử

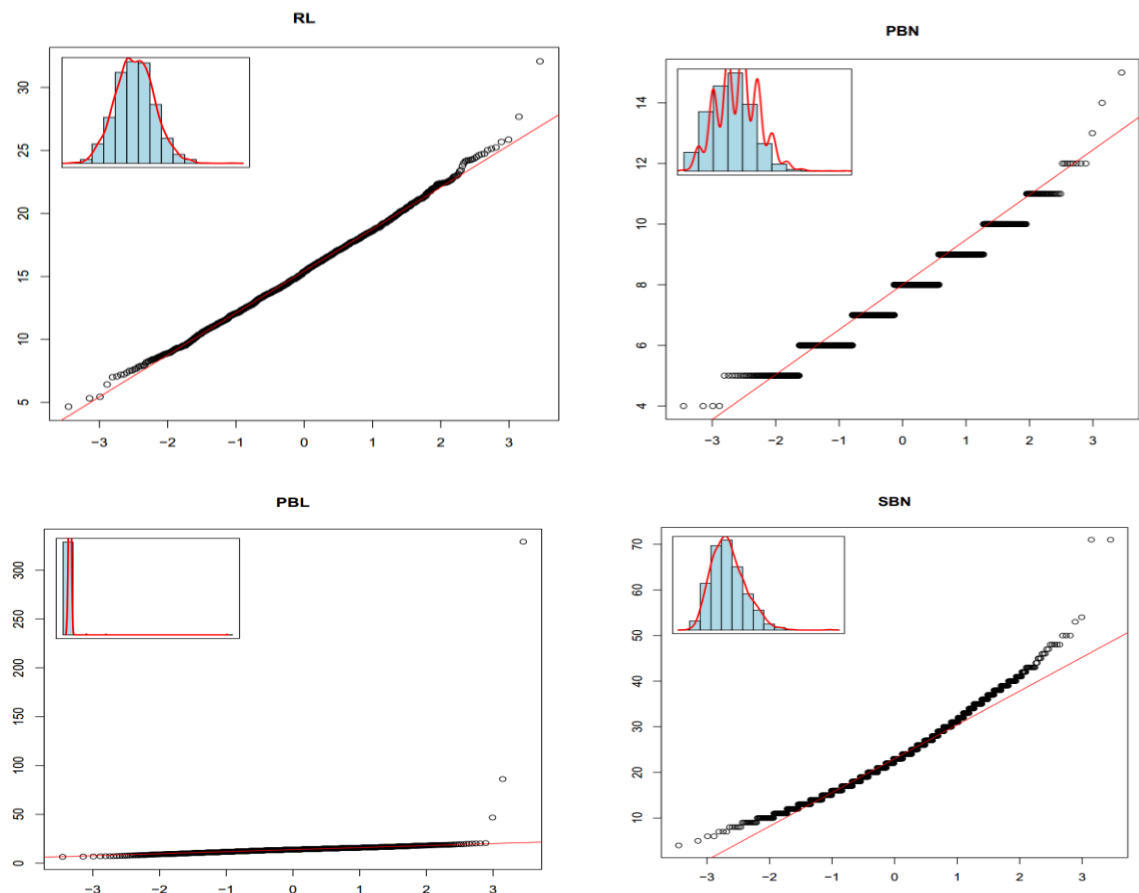
(Vụ xuân năm 2019 tại xã Đông La, Hoài Đức, Hà Nội)

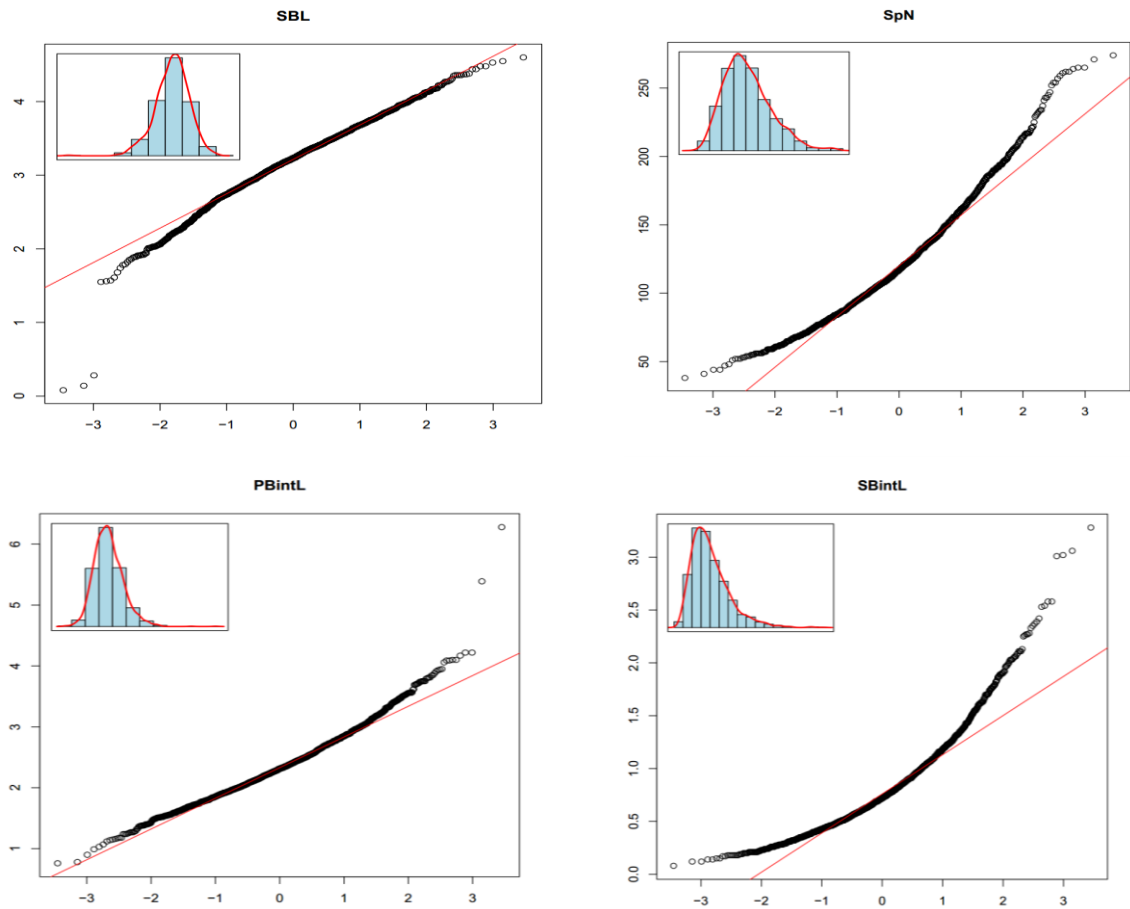
STT	Haplotype				
	Chỉ tiêu theo dõi	G6 (H1)	G189 (H2)	F3_H1	F3_H2
1	Chiều dài trục bông (cm)	20,1±2,3	15,6±3,5	15,7±3,8	15,9±3,4
2	Đường kính trục bông (cm)	0,2±0,04	0,2±0,07	0,2±0,06	0,2±0,06
3	Số đốt/trục chính (đốt)	8,1±1,1	5,9±0,7	7,0±1,3	7,5±1,4
4	Số gié cấp một (gié)	9,7±0,9	6,2±0,6	7,7±1,5	7,9±1,4
5	Chiều dài trung bình gié cấp một (cm)	13,8±0,8	15,9±1,5	13,8±1,4	14,1±2,3
6	Khoảng cách giữa các gié cấp một (cm)	2,3±0,2	3,0±0,6	2,4±0,4	2,4±0,5
7	Số gié cấp hai/bông (gié)	18,6±2,3	26,0±6,0	20,0±7,1	23,4±7,8
8	Chiều dài trung bình gié cấp hai (cm)	3,2±0,4	3,2±0,4	3,2±0,5	3,2±0,4
9	Khoảng cách giữa các gié cấp hai (cm)	1,6±0,5	0,6±0,2	1,0±0,6	0,9±0,4
10	Số gié cấp ba/bông (gié)	0,1±0,3	3,0±3,4	0,3±0,7	0,6±1,4
11	Số hạt/bông (hạt)	108,0±11,8	140,0±39,2	104,4±32,3	119,0±34,2
12	Chiều cao cây (cm)	134,6±3,5	108,8±13,4	123,5±11,7	126,2±13,1
13	Số nhánh/cây (nhánh)	11,5±3,0	6,3±1,6	10,1±2,8	9,1±2,7
14	Số nhánh hữu hiệu/cây (nhánh)	11,5±2,3	6,0±1,2	9,9±2,7	8,9±2,6

STT	Haplotype				
	Chỉ tiêu theo dõi	G6 (H1)	G189 (H2)	F3_H1	F3_H2
15	Ngày bông đầu tiên ra hoa (ngày)	86,5±2,3	70,3±1,3	75,3±10,1	71,8±11,2
16	Ngày 50% số bông ra hoa (ngày)	93,5±1,7	76,0±4,3	84,8±7,1	82,3±8,9

Các số liệu đo đếm trên được xử lý và phân tích bằng phần mềm thống kê R của R Core Team, 2019 [92].

Đánh giá sự phân bố kiểu hình của các thành phần cấu trúc bông cho thấy, các tính trạng phân bố theo quy luật phân bố chuẩn hoặc gần phân bố chuẩn vì các giá trị quan sát nằm gần và nằm trên đường kỳ vọng (màu đỏ) (hình 3.16). Điều đó cho thấy kết quả đo đếm các chỉ tiêu có độ tin cậy cao.

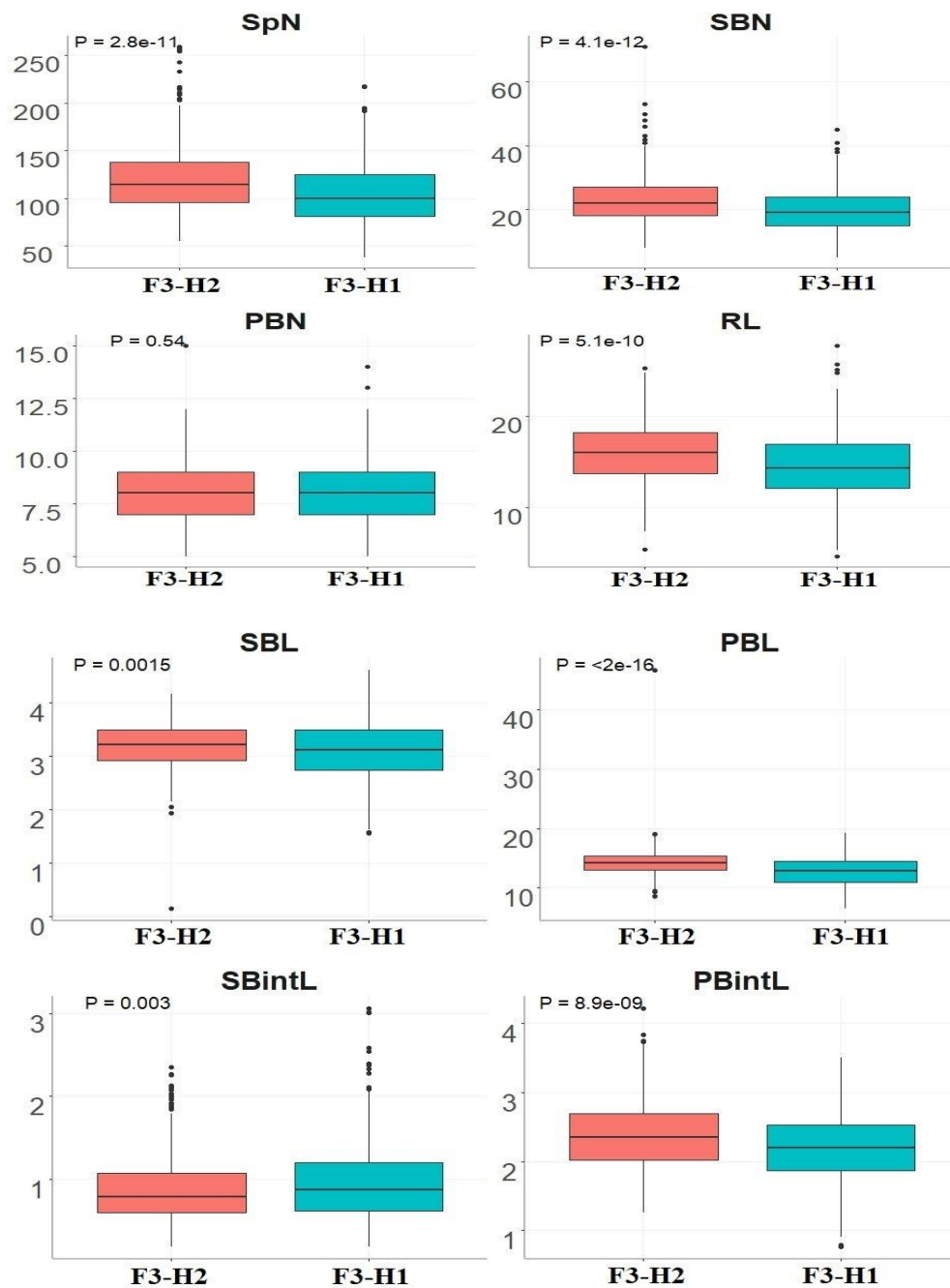




Hình 3.16. Biểu đồ Q-Q plot về sự phân bố của các tính trạng cấu trúc bông lúa

(*RL: Chiều dài bông; PBN: Số gié cấp một/bông; PBL: Chiều dài gié cấp một; SBN: Số gié cấp hai/bông; SBL: Chiều dài gié cấp hai; SpN: Số hạt/bông; PBintL: Khoảng cách giữa các gié cấp một; SBintL: Khoảng cách giữa các gié cấp hai*)

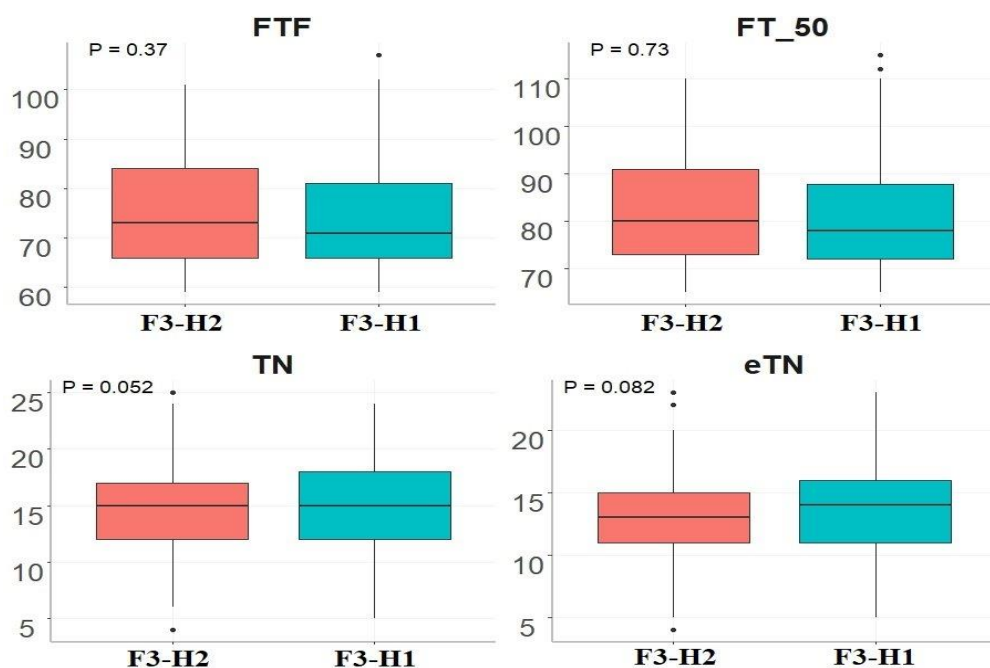
So sánh tám tính trạng cấu trúc bông và bốn tính trạng liên quan đến năng suất của quần thể F_3 đồng hợp tử được thể hiện qua hình 3.17 và hình 3.18.



Hình 3.17. Biểu đồ hộp so sánh 8 tính trạng cấu trúc bông của quần thể F₃ đồng hợp tử mang QTL9 của mẹ (F3_H1) và bố (F3_H2)

(Ghi chú: RL: chiều dài trục bông; PBN: số gié cấp một/bông; SBN: số gié cấp hai/bông; TBN: số gié cấp ba/bông; SpN: số hạt/bông; PBL: chiều dài gié cấp một; SBL: chiều dài gié cấp hai; TBL: chiều dài gié cấp ba; PBintL: khoảng cách giữa các gié cấp một; SBintL: khoảng cách giữa các gié cấp hai)

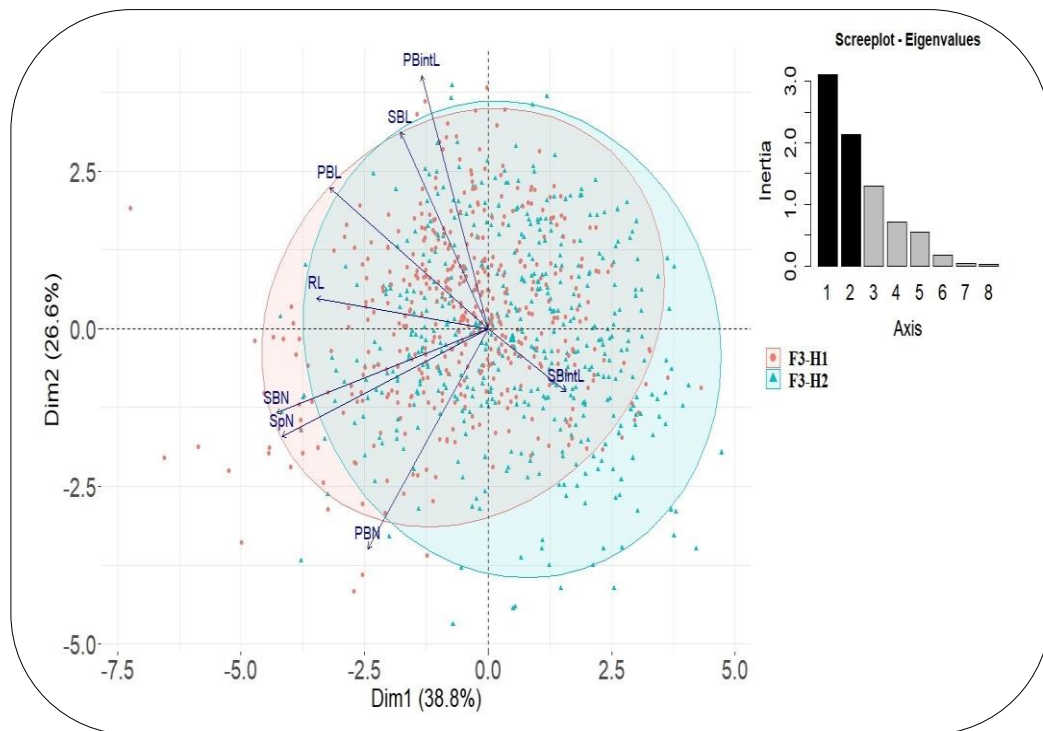
Hình 3.17, so sánh tám tính trạng cấu trúc bông giữa hai haplotype trong quần thể F₃ cho thấy các tính trạng số hạt/bông (SpN), số gié cấp hai (SBN), chiều dài bông (RL), khoảng cách giữa các gié cấp một (PBintL), chiều dài gié cấp hai (SBL) ở H2 cao hơn H1 có ý nghĩa thống kê ($p < 0,005$), ngược lại tính trạng khoảng cách giữa các gié cấp hai (SBintL) ở H2 thấp hơn ở H1 với $p = 0,003$. Kết quả cũng cho thấy không có sự khác biệt về số gié cấp một (PBN) giữa hai haplotype ($p > 0,005$).



Hình 3.18. Biểu đồ hộp so sánh bốn tính trạng liên quan đến năng suất của quần thể F₃ đồng hợp tử mang QTL9 của mẹ (F3_H1) và bố (F3_H2) (Ghi chú: TN: Số nhánh/cây; eTN: số nhánh hữu hiệu/cây; FTF: ngày ra hoa bông đầu tiên; FT_50: ngày ra hoa của 50% số bông)

Hình 3.18, so sánh bốn tính trạng liên quan đến năng suất gồm số nhánh/cây (TN), số nhánh hữu hiệu/cây (eTN), ngày bông đầu tiên ra hoa (FTF) và ngày có 50% số bông ra hoa (FT-50) cho thấy cả bốn tính trạng này đều có khác biệt không có ý nghĩa thống kê ($p > 0,05$). Hay nói cách khác, các tính trạng này không có sự khác biệt giữa hai haplotype.

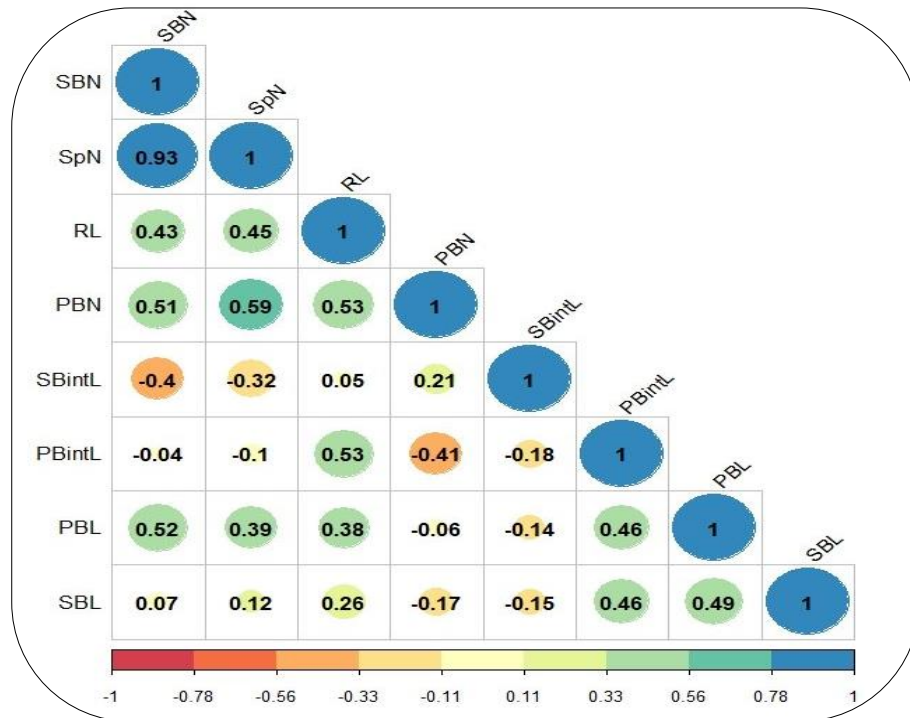
Tương tự như vậy, khi phân tích tương quan các thành phần chính PCA (hình 3.19) cho thấy các tính trạng liên quan đến cấu trúc bông chia thành hai nhóm: nhóm liên quan đến số lượng (số gié cấp một - PBN, số gié cấp hai - SBN và số hạt/bông - SpN) và nhóm liên quan đến chiều dài (chiều dài bông - PL, chiều dài gié cấp một - PBL, chiều dài gié cấp hai - SBL, khoảng cách giữa các gié cấp một - PBintL). Trong đó, số hạt/bông (SpN) có mối liên hệ chặt chẽ với số gié cấp hai/bông (SBN) mối liên hệ thấp hơn với số gié cấp một/bông (PBN). Ngược lại, khoảng cách giữa các gié cấp hai (SBintL) không liên quan đến các tính trạng còn lại (Hình 3.19).



Hình 3.19. Biểu đồ phân tích tương quan thành phần chính PCA

(Biểu đồ hình cột bên phải biểu diễn toàn bộ dữ liệu thông qua 8 cột trên trục hoành Axis, trong đó hai thành phần chính Axis 1 và Axis 2 (màu đen) phản ánh hơn 50% bộ dữ liệu. Biểu đồ hệ tọa độ bên trái thể hiện ảnh xạ dữ liệu hai thành phần chính Axis 1 (Dim1) và Axis 2 (Dim2). Mỗi chấm là hình ảnh ảnh xạ của biến lên hệ tọa độ với 2 thành phần chính (Dim1, Dim2), thành phần thứ nhất phản ánh 38,8% bộ dữ liệu, thành phần thứ 2 phản ánh 26,6% bộ dữ liệu).

Mối tương quan giữa các tính trạng được thể hiện rõ hơn qua biểu đồ phân tích tương quan corrplot (hình 3.20).



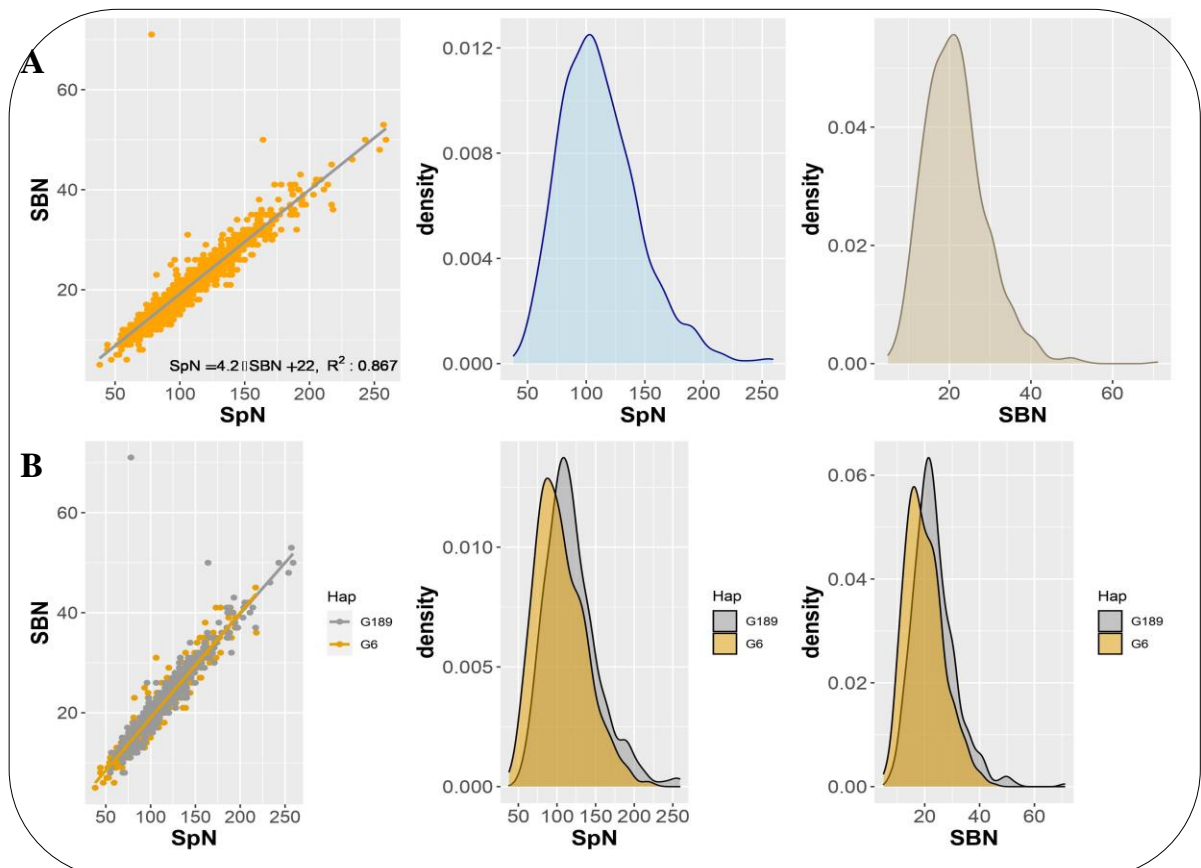
Hình 3.20. Biểu đồ mối tương quan Corrplot giữa các tính trạng cấu trúc bông ở quần thể F₃

(Ghi chú: RL: chiều dài trục bông; PBN: số gié sơ cấp/bông; SBN: số gié thứ cấp/bông; TBN: số gié tam cấp/bông; SpN: số hạt/bông; PBL: chiều dài gié sơ cấp; SBL: chiều dài gié thứ cấp; TBL: chiều dài gié tam cấp; PBintL: khoảng cách giữa các gié sơ cấp; SBintL: khoảng cách giữa các gié thứ cấp; TN: Số nhánh/cây; eTN: số nhánh hữu hiệu/cây)

Phân tích mối tương quan giữa các tính trạng cấu trúc bông (hình 3.20) cho thấy số gié cấp hai/bông (SBN) và số hạt/bông (SpN) có mối tương quan chặt chẽ với nhau với hệ số tương quan 0,93. Trong khi đó, các tính trạng số gié cấp một/bông (PBN) và số gié cấp hai/bông (SBN); số gié cấp một/bông (PBN) và số hạt/bông (SpN) hay chiều dài các gié cấp một (PBL) và số hạt/bông (SpN) có mối tương quan vừa với hệ số tương quan từ 0,51 - 0,59. Các tính trạng có mối tương quan thấp như chiều dài bông (RL)

và số gié cấp hai (SBN); số hạt/bông (SpN), chiều dài các gié cấp một (PBL) và số hạt/bông (SpN) với hệ số tương quan 0,39 - 0,45. Ngược lại, có một số tính trạng không có mối tương quan với nhau như số gié cấp hai/bông (SBN), số hạt/bông (SpN) và khoảng cách giữa các gié cấp hai (SBintL); khoảng cách giữa các gié cấp một (PBintL) với hệ số tương quan âm (-0,4 đến -0,04). Các kết quả này hoàn toàn tương đồng với kết quả phân tích GWAS đã được công bố trước đây trên tập đoàn lúa bản địa Việt Nam của Tạ Kim Nhung và cộng sự năm 2018 [102]. Điều đó cho thấy, các kết quả nghiên cứu của đề tài là đáng tin cậy.

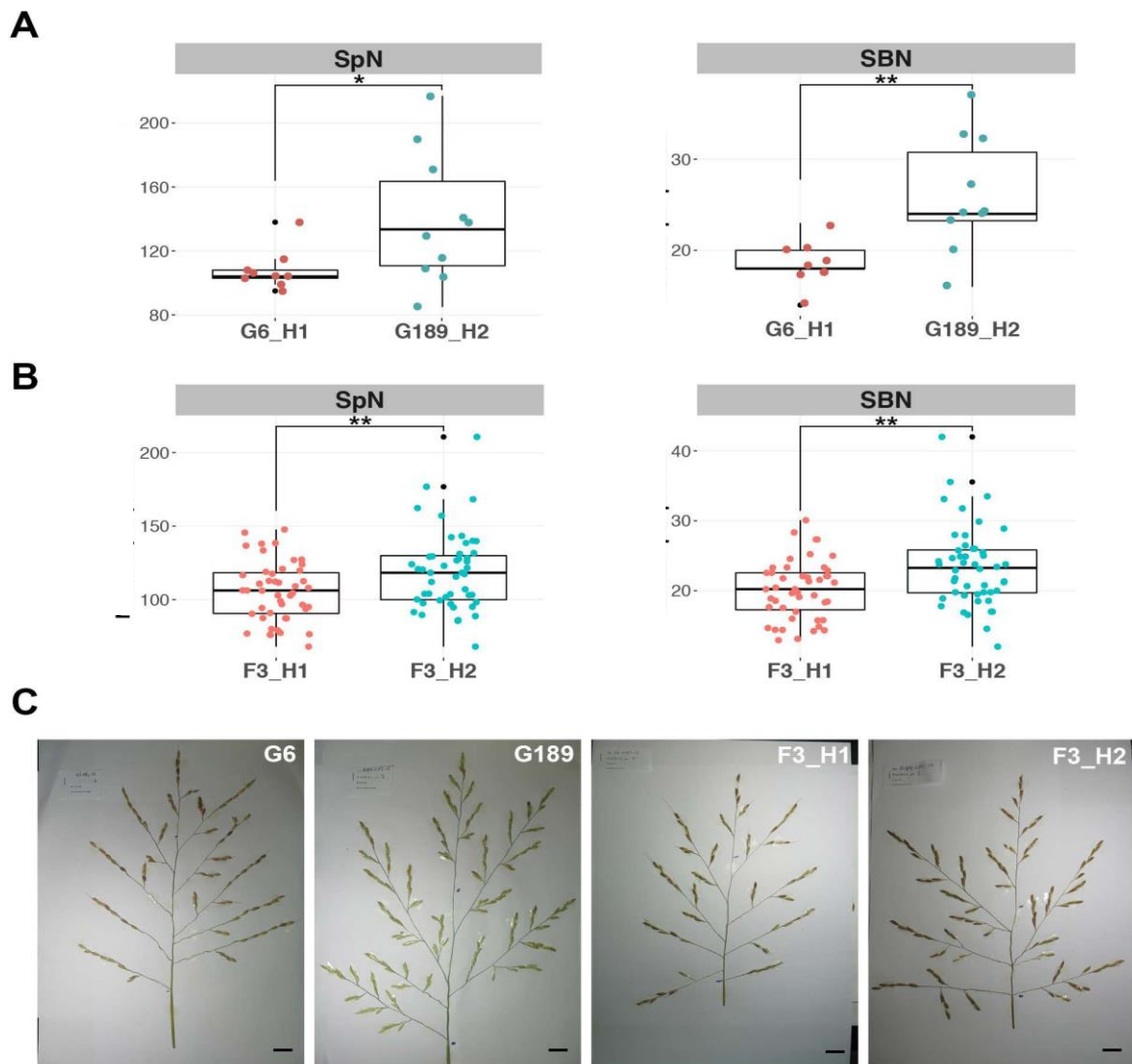
Mô hình hồi quy tuyến tính của các tính trạng số gié cấp hai/bông (SBN) và số hạt/bông (SpN) của quần thể F_3 thể hiện qua hình 3.21.



Hình 3.21. Mô hình hồi quy tuyến tính đơn biến giữa hai tính trạng số gié cấp hai/bông (SBN) và số hạt/bông (SpN) (A) và tương quan tuyến tính giữa các tính trạng SpN và SBN trong hai quần thể con F_3 thuộc hai haplotype (B)

Theo tác giả Nguyễn Văn Cường, 2014 [1] trong phân tích QTL, một giá trị cần được chú ý là giá trị R^2 hay PVE (Phenotypic variance explained - phần trăm giải thích sự khác biệt kiểu hình tại vị trí QTL đó). Trong nghiên cứu của Ashikari và cộng sự, 2005 [14] nghiên cứu lập bản đồ QTL các tính trạng liên quan đến năng suất lúa thông qua các quần thể lai giữa hai giống lúa có năng suất khác nhau, thuộc hai nhóm *Japonica* và *Indica* cũng đã xác định được QTL làm tăng số hạt/bông *Gn1* có giá trị R^2 cao nhất 44% cho thấy có thể dễ dàng phân biệt được các nhóm kiểu hình khác nhau (QTL này làm tăng số lượng hạt tới 92 hạt). Do đó, xác suất có gen trong vùng QTL này ảnh hưởng đến tính trạng là rất cao. Như vậy, qua mô hình hồi quy tuyến tính đơn (hình 3.21A) có thể chỉ ra rằng mức độ đa dạng của số gié cấp hai trên bông (SBN) có thể lý giải cho 86,7% sự đa dạng của số hạt trên bông (SpN).

QTL9 là một QTL mới phát hiện liên quan đến hai tính trạng số gié cấp hai/bông (SBN) và số hạt/bông (SpN) thu được từ nghiên cứu phân tích GWAS của Tạ Kim Nhung và các cộng sự (2018) trên tập đoàn 159 giống lúa bản địa Việt Nam, trong đó có hai giống G6 và G189. Trong đó, giống G6 có cấu trúc bông nhỏ, mang QTL9 haplotype 1; giống G189 có cấu trúc bông to mang QTL9 haplotype 2. Quần thể F_3 được tạo ra từ cặp lai G6 và G189. Vì vậy, để đánh giá được vai trò của QTL9 liên quan đến cấu trúc bông lúa, cụ thể là hai tính trạng số gié cấp hai/bông (SBN) và số hạt/bông (SpN), tiến hành so sánh hai tính trạng này giữa hai haplotype ở quần thể F_3 đồng hợp tử và các cây bố mẹ G6 và G189. Kết quả được thể hiện qua hình 3.22.



Hình 3.22. So sánh cấu trúc bông của hai haplotype

(A) Biểu đồ hình hộp với các chấm riêng lẻ cho biết số lượng hạt (SpN) và số gié cấp hai (SBN) trên mỗi bông ở hai nhóm bố mẹ G6 từ haplotype H1 và G189 từ haplotype H2. (B) Biểu đồ hình hộp với các chấm riêng lẻ biểu thị số lượng hạt (SpN) và số gié cấp hai (SBN) trên mỗi bông của quần thể F_3 từ quần thể F_1 của cặp lai G6xG189 có haplotype H1 (F3_H1) hoặc haplotype H2 (F3_H2) trong vùng QTL9. (C) Bông lúa của giống bố mẹ G6 và G189, một dòng F3_H1 và một dòng F3_H2 chụp tại Viện Di truyền Nông nghiệp năm 2019. Các chấm riêng lẻ trong hộp tương ứng với giá trị trung bình từ 3 bông chính trên mỗi cây. Thanh chia độ: 2 cm. Ý nghĩa thống kê (giá trị p kiểm định) giữa hai dòng hoặc bố mẹ đối với hai tính trạng hình thái bông được biểu thị như sau: * nếu giá trị $p < 0,05$, ** nếu $p < 0,01$, *** nếu $p < 0,001$.

Hình 3.22A so sánh số hạt/bông và số gié cấp hai/bông giữa hai haplotype ở quần thể bố mẹ. Kết quả cho thấy:

Số hạt/bông ở quần thể mẹ (G6_H1) chủ yếu nằm trong khoảng 104 - 110 hạt/bông, một số cây có số hạt thấp dưới 100 hạt và cao trên 140 hạt/bông. Trong khi đó, số hạt/bông ở quần thể bố (G189_H2) cao hơn hẳn quần thể mẹ (G6_H1) có ý nghĩa thống kê với $p < 0,05$; với số hạt trung bình từ 110 - 160 hạt/bông, một số cây đạt trên 200 hạt/bông.

Tương tự như vậy, đối với số gié cấp hai/bông (SBN) ở quần thể bố cao hơn quần thể mẹ có ý nghĩa thống kê với $p < 0,01$ (hình 3.22A). Ở quần thể mẹ (G6_H1) có số gié trung bình/bông từ 18 - 20 gié; trong khi ở quần thể bố (G189_H2) đạt từ 24 - 32 gié cấp hai/bông.

Tiếp đến, hình 3.22B so sánh số hạt/bông và số gié cấp hai/bông giữa hai haplotype ở quần thể F₃ đồng hợp tử. Quần thể này được tạo ra từ quần thể F₁ của cặp lai G6 và G189 tự thụ phấn tạo quần thể F₂, sau đó các cây F₂ đồng hợp tử mang QTL9 của bố hoặc của mẹ được chọn lọc bằng chỉ thị phân tử CAPS (kết quả ở nội dung trên) tiếp tục cho tự thụ phấn tạo quần thể F₃ đồng hợp tử. Kết quả hình 3.22B cho thấy cả hai chỉ tiêu số hạt/bông (SpN) và số gié cấp hai/bông (SBN) ở quần thể F3_H2 đều cao hơn quần thể F3_H1, sự khác biệt này có ý nghĩa thống kê giữa hai haplotype với $p < 0,01$. Số hạt/bông trung bình ở quần thể F3_H1 khoảng 104 hạt; trong khi đó ở quần thể F3_H2 khoảng 118 hạt. Số gié cấp hai/bông ở quần thể F3_H1 khoảng 19 gié và ở quần thể F3_H2 khoảng 23 gié. Sự khác biệt giữa hai haplotype ở quần thể F₃ đồng hợp tử cũng tương đồng với các giống bố mẹ.

Đặc biệt, số gié cấp hai/bông (SBN) và số hạt/bông (SpN) ở quần thể F₃ đồng hợp tử mang QTL9 cây bố (F3_H2) cao hơn số gié cấp hai/bông (SBN) và số hạt/bông (SpN) ở quần thể F₃ đồng hợp tử mang QTL9 cây mẹ (F3_H1). Kết quả nghiên cứu này phù hợp với kết quả phân tích GWAS trước

đây của Tạ Kim Nhung và cộng sự năm 2018 [102], QTL9 liên kết với hai tính trạng số gié cấp hai/bông và số hạt/bông. Điều này khẳng định sự quan trọng của vùng QTL9 đối với sự phân nhánh của bông, những cây mang QTL9_H2 có xu hướng bông to, nhiều hạt hơn những cây mang QTL9_H1.

KẾT LUẬN VÀ KIẾN NGHỊ

1. Kết luận

1.1. Phân tích haplotype vùng QTL9 đã xác định được 04 giống lúa làm bố mẹ thuộc hai haplotype H1, H2 có cấu trúc bông khác biệt (bông to và bông nhỏ). Trong đó giống có cấu trúc bông to (G189, G205) thuộc H2 làm cây bố và giống có cấu trúc bông nhỏ (G6, G19) thuộc H1 làm cây mẹ và lai tạo được 04 quần thể lai F_1 từ 04 giống lúa đã chọn lọc. Chọn được 7/12 chỉ thị SSR cho đa hình giữa các giống bố mẹ sử dụng để chọn lọc cây F_1 . Đã xác định được: 12/20 cây lai F_1 của cặp G6 x G189 nhờ 3 chỉ thị RM320, RM180, RM491; 11/20 cây lai F_1 của cặp G6 x G205 nhờ 2 chỉ thị RM320, RM180; 17/20 cây lai F_1 của cặp G19 x G189 nhờ 3 chỉ thị RM289, RM592, RM204; 11/20 cây lai F_1 của cặp G19 x G205 nhờ 3 chỉ thị RM289, RM592, RM204.

1.2. Đã phát triển được được 3 chỉ thị phân tử CAPS bao phủ vùng QTL9: CAPS 1, CAPS 5, CAPS 11 cho đa hình giữa các giống bố mẹ để chọn lọc các cây F_2 đồng hợp tử.

1.3. Đã tạo được quần thể F_2 gồm 275 cá thể bằng ba chỉ thị phân tử CAPS; chọn lọc được 74 cây F_2 đồng hợp tử thuộc H1, 66 cây F_2 đồng hợp tử thuộc H2, chiếm tỷ lệ 50,9% bằng ba chỉ thị phân tử CAPS. Các cây F_2 đồng hợp tử được phát triển tạo quần thể F_3_H1 và F_3_H2 để đánh giá cấu trúc bông.

1.4. Phân tích kiểu hình cấu trúc bông của hai quần thể F_3 gồm 1.568 cá thể đồng hợp tử mang kiểu gen QTL9 của bố hoặc của mẹ đã xác định được QTL9 có tham gia vào việc quy định kiểu hình cấu trúc bông, đặc biệt là hai tính trạng số gié cấp hai/bông và số hạt/bông. Những cá thể mang QTL9 thuộc H1 có xu hướng mang cấu trúc bông nhỏ và những cá thể mang QTL9 thuộc H2 có xu hướng mang cấu trúc bông to.

2. Kiến nghị

Đưa QTL9 và bộ chỉ thị phân tử CAPS ứng dụng vào chương trình chọn tạo giống lúa năng suất cao ở Việt Nam.

DANH MỤC

CÔNG TRÌNH ĐÃ CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN

1. Vũ Thị Nhiên, Tạ Kim Nhung, Stefan Jouanic, Lê Hùng Lĩnh, Phạm Xuân Hội, Trần Khánh Vân, Trần Vũ Hằng, **Phạm Thị Mai**, Lê Thị Như, Không Ngân Giang (2018), “Tạo quần thể lai F1 làm vật liệu khởi đầu để đánh giá vai trò của QTL9 liên quan đến các tính trạng năng suất của tập đoàn lúa bản địa Việt Nam”, *Tạp chí Khoa học Công nghệ Nông nghiệp Việt Nam*, số 11 (96) 2018, tr.25-32.
2. Stefan Jouanic, **Phạm Thị Mai**, Lê Thị Như, Phạm Xuân Hội, Không Ngân Giang (2020), Ứng dụng công nghệ gene capture để xác định SNP trong vùng *QTL9* liên quan đến cấu trúc bông lúa, *Hội nghị Công nghệ sinh học toàn quốc 2020*, tr.162-167.
3. **Phạm Thị Mai**, Lê Thị Như, Stefan Jouannic, Không Ngân Giang (2021), Chọn lọc các cá thể F₂ đồng hợp tử mang QTL9 liên quan đến cấu trúc bông lúa bằng chỉ thị phân tử CAPS, *Tạp chí Khoa học và Công nghệ tập 63, số 2, tháng 2/2021*, tr.55-59.
4. Giang Ngan Khong, Nhu Thi Le, **Mai Thi Pham**, Helene Adam, Carole Gauron, Hoa Quang Le, Dung Tien Pham, Kelly Colonges, Xuan Hoi Pham, Vinh Nang Do, Michel Lebrun, Stefan Jouannic (2021), A cluster of Ankyrin and Ankyrin-TPR repeat genes is associated with panicle branching diversity in rice, *PLOS Genetics* tháng 6/2021, <http://doi.org/10.1371/journal.pgen.1009594>

TÀI LIỆU THAM KHẢO

A. Tiếng Việt

1. Nguyễn Văn Cường (2014), “Tìm hiểu và phân tích tính trạng số lượng ở thực vật”, *Vietnam Journal of Science*. <http://www.vjsonline.org/research-review/>
2. Phạm Văn Cường, Tăng Thị Hạnh, Vũ Văn Liết, Nguyễn Thiện Huyền, Nguyễn Hữu Tề (2015), *Giáo trình cây lúa*. Nhà xuất bản Đại học Nông Nghiệp.
3. Nguyễn Đình Giao, Nguyễn Thiện Huyền, Nguyễn Hữu Tề, Hà Công Vượng (2001), *Giáo trình cây lương thực tập 1: Cây Lúa*. Nhà xuất bản Nông nghiệp, Hà Nội.
4. Lê Huy Hàm, Trần Đăng Khánh (2015), *Ứng dụng chỉ thị phân tử trong chọn tạo giống lúa*. Nhà xuất bản Nông nghiệp, Hà Nội, tr.219-236.
5. Nguyễn Văn Hoan (2006), *Cẩm nang cây lúa*. Nhà xuất bản Lao động, Hà Nội.
6. Phạm Xuân Hội (chủ biên), Trần Đăng Khánh, Khuất Hữu Trung, Võ Thị Minh Tuyên, Lưu Minh Cúc, Không Ngân Giang, Phùng Thị Phương Nhung, Nguyễn Thị Minh Nguyệt, Nguyễn Duy Phương (2019), *Công nghệ sinh học và triển vọng ứng dụng trong chọn tạo giống lúa ở Việt Nam*. MS: 284-KHTN-2019. Nhà xuất bản Đại học Quốc Gia, Hà Nội, tr.69-82.
7. Lê Hùng Lĩnh, Lưu Minh Cúc (2016), *Ứng dụng chỉ thị phân tử tích hợp gen/QTLs trong cải tiến giống lúa*. Nhà xuất bản Nông nghiệp, Hà Nội.
8. Nguyễn Văn Tuấn (2018), *Phân tích dữ liệu với R hỏi và đáp*, Nhà xuất bản tổng hợp Thành phố Hồ Chí Minh.

B. Tiếng Anh

9. Abbai R., Singh V. K., Nachimuthu V. V., Sinha P., Selvaraj R., Vipparla A. K., Singh A. K., Singh U. M., Varshney R. K., & Kumar A. (2019), "Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains", *Plant Biotechnology Journal*, 17(8), 1612–1622.
10. Akagi H., Yokozeki Y., Inagaki A., & Fujimura T. (1996), "Microsatellite DNA markers for rice chromosomes", *Theoretical and Applied Genetics*, 93(7), 1071–1077.
11. Aktas C. (2020), *Haplotypes: Manipulating DNA Sequences and Estimating Unambiguous Haplotype Network with Statistical Parsimony* (1.1.2) [Computer software]. <https://CRAN.R-project.org/package=haplotypes>.
12. AL-Tam F., Adam H., Anjos A. dos, Lorieux M., Larmande P., Ghesquière A., Jouannic S., & Shahbazkia H. R. (2013), "P-TRAP: A Panicle Trait Phenotyping tool", *BMC Plant Biology*, 13(1), 122.
13. Arora V., Ghosh M. K., Pal S., & Gangopadhyay G. (2017), "Allele specific CAPS marker development and characterization of chalcone synthase gene in Indian mulberry (*Morus spp.*, family *Moraceae*)", *PLoS ONE*, 12(6), e0179189.
14. Ashikari M., Sakakibara H., Lin S., Yamamoto T., Takashi T., Nishimura A., Angeles E. R., Qian Q., Kitano H., & Matsuoka M. (2005), "Cytokinin oxidase regulates rice grain production", *Science (New York, N.Y.)*, 309(5735), 741–745.
15. Babu B. K., Mathur R. K., Kumar P. N., Ramajayam D., Ravichandran G., Venu M. V. B., & Babu S. S. (2017), "Development, identification and validation of CAPS marker for SHELL trait which governs dura, pisifera

- and tenera fruit forms in oil palm (*Elaeis guineensis* Jacq.)", *PLOS ONE*, 12(2), e0171933.
16. Bai S., Hong J., Li L., Su S., Li Z., Wang W., Zhang F., Liang W., & Zhang D. (2021), "Dissection of the Genetic Basis of Rice Panicle Architecture Using a Genome-wide Association Study", *Rice*, 14(1), 77.
 17. Bai X., Luo L., Yan W., Kovi M. R., Zhan W., & Xing Y. (2010), "Genetic dissection of rice grain shape using a recombinant inbred line population derived from two contrasting parents and fine mapping a pleiotropic quantitative trait locus qGL7", *BMC Genetics*, 11(1), 16.
 18. Bai X., Wu B., & Xing Y. (2012), "Yield-related QTLs and their applications in rice genetic improvement", *Journal of Integrative Plant Biology*, 54(5), 300–311.
 19. Bai X., Zhao H., Huang Y., Xie W., Han Z., Zhang B., Guo Z., Yang L., Dong H., Xue W., Li G., Hu G., Hu Y., & Xing Y. (2016), "Genome-Wide Association Analysis Reveals Different Genetic Control in Panicle Architecture Between and Rice", *The Plant Genome*, 9(2), 1-10.
 20. Balkunde S., Le H.-L., Lee H.-S., Kim D.-M., Kang J.-W., & Ahn S.-N. (2013), "Fine mapping of a QTL for the number of spikelets per panicle by using near-isogenic lines derived from an interspecific cross between *Oryza sativa* and *Oryza minuta*", *Plant Breeding*, 132(1), 70–76.
 21. Bao J. S., Wu Y. R., Hu B., Wu P., Cui H. R., & Shu Q. Y. (2002), "QTL for rice grain quality based on a DH population derived from parents with similar apparent amylose content", *Euphytica*, 128(3), 317–324.
 22. Bargsten J. W., Nap J.-P., Sanchez-Perez G. F., & van Dijk A. D. J. (2014), "Prioritization of candidate genes in QTL regions based on associations between traits and biological processes", *BMC Plant Biology*, 14, 330.

23. Benner C., Havulinna A. S., Järvelin M.-R., Salomaa V., Ripatti S., & Pirinen M. (2017), "Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies", *American Journal of Human Genetics*, 101(4), 539–551.
24. Bevan M. W., & Uauy C. (2013) "Genomics reveals new landscapes for crop improvement", *Genome Biology*, 14(6), 206.
25. Bevan M. W., Uauy C., Wulff B. B. H., Zhou J., Krasileva K., & Clark M. D. (2017), "Genomic innovation for crop improvement", *Nature*, 543(7645), 346–354.
26. Bhat R., Singh A. K., Salgotra R. K., Sharma M., Mushtaq M., Bagati S., Hangloo S., & Singh A. (2019), "Detection of QTL for panicle architecture in F2 population of rice", *Journal of Genetics*, 98(2), 50.
27. Bhat J. A., Yu D., Bohra A., Ganie S. A., & Varshney R. K. (2021), "Features and applications of haplotypes in crop breeding", *Communications Biology*, 4(1), 1–12.
28. Bogacki P., Peck D. M., Nair R. M., Howie J., & Oldach K. H. (2013), "Genetic analysis of tolerance to Boron toxicity in the legume *Medicago truncatula*", *BMC Plant Biology*, 13, 54.
29. Borgström E., Lundin S., & Lundeberg J. (2011), "Large scale library generation for high throughput sequencing", *PloS One*, 6(4), e19119.
30. Chen X., Temnykh S., Xu Y., Cho Y. G., & McCouch S. R. (1997), "Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.)", *Theoretical and Applied Genetics*, 95(4), 553–567.
31. Cheng A., Ismail I., Osman M., & Hashim H. (2012), "Simple and Rapid Molecular Techniques for Identification of Amylose Levels in Rice

- Varieties", *International Journal of Molecular Sciences*, 13(5), 6156–6166.
32. Clark M., Chen R., Lam H., Karczewski K., Chen R., Euskirchen G., Butte A., & Snyder M. (2011), "Performance Comparison of Exome Dna Sequencing Technologies", *Nature Biotechnology*, 29, 908–914.
 33. CRAN - Package *corrplot*. (2020) from <https://cran.r-project.org/web/packages/corrplot/index.html>
 34. Crowell S., Korniliev P., Falcão A., Ismail A., Gregorio G., Mezey J., & McCouch S. (2016), "Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters", *Nature Communications*, 7(1), 10527.
 35. Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., DePristo M. A., Handsaker R. E., Lunter G., Marth G. T., Sherry S. T., McVean G., & Durbin R. (2011), "The variant call format and VCFtools", *Bioinformatics*, 27(15), 2156–2158.
 36. DePristo M. A., Banks E., Poplin R., Garimella K. V., Maguire J. R., Hartl C., Philippakis A. A., del Angel G., Rivas M. A., Hanna M., McKenna A., Fennell T. J., Kernytsky A. M., Sivachenko A. Y., Cibulskis K., Gabriel S. B., Altshuler D., & Daly M. J. (2011), "A framework for variation discovery and genotyping using next-generation DNA sequencing data", *Nature Genetics*, 43(5), 491–498.
 37. Dereeper A., Nicolas S., Le Cunff L., Bacilieri R., Doligez A., Peros J.-P., Ruiz M., & This P. (2011), "SNiPlay: A web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects", *BMC Bioinformatics*, 12, 134.
 38. Dewi E. A., Michael D., & Audrey D., Hélène A., Delphine L. & Tanguy L. (2016), "Rice panicle plasticity in Near Isogenic Lines carrying a QTL

- for larger panicle is genotype and environment dependent", *Rice (New York, N.Y.)*, 9(1), 28
39. Donde R., Mohapatra S., Baksh S. K. Y., Padhy B., Mukherjee M., Roy S., Chattopadhyay K., Anandan A., Swain P., Sahoo K. K., Singh O. N., Behera L., & Dash S. K. (2020), "Identification of QTLs for high grain yield and component traits in new plant types of rice", *PLOS ONE*, 15(7), e0227785.
 40. Doyle J. (1991), "DNA Protocols for Plants", In G. M. Hewitt, A. W. B. Johnston, & J. P. W. Young (Eds.), *Molecular Techniques in Taxonomy* (pp. 283–293).
 41. Fan C., Xing Y., Mao H., Lu T., Han B., Xu C., Li X., & Zhang Q. (2006), "GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein", *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 112(6), 1164–1171.
 42. Flint-Garcia S. A., Thornsberry J. M., & Buckler E. S. (2003), "Structure of linkage disequilibrium in plants", *Annual Review of Plant Biology*, 54, 357–374.
 43. Hasina B., Jennifer E. S., & Antonio L., Teresita B., Glenn G., Jose H., Parminder V., Bertrand C., Susan R. McCouc. (2015), "Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*)", *PloS One*, 10(3), e0119873
 44. Hazarika T. K., Hazarika B. N., & Shukla A. C. (2014), "Genetic variability and phylogenetic relationships studies of genus *Citrus* L. with the application of molecular markers", *Genetic Resources and Crop Evolution*, 61(8), 1441–1454.

45. Heubl G. (2010), "New aspects of DNA-based authentication of Chinese medicinal plants by molecular biological techniques", *Planta Medica*, 76(17), 1963–1974
46. Hill C. B., Wong D., Tibbits J., Forrest K., Hayden M., Zhang X.-Q., Westcott S., Angessa T. T., & Li C. (2019), "Targeted enrichment by solution-based hybrid capture to identify genetic sequence variants in barley", *Scientific Data*, 6(1), 1–8.
47. Hittalmani S., Shashidhar H. E., Bagali P. G., Huang N., Sidhu J. S., Singh V. P., & Khush G. S. (2002), "Molecular mapping of quantitative trait loci for plant growth, yield and yield related traits across three diverse locations in a doubled haploid rice population", *Euphytica*, 125(2), 207–214.
48. Hoang G. T., Van Dinh L., Nguyen T. T., Ta N. K., Gathignol F., Mai C. D., Jouannic S., Tran K. D., Khuat T. H., Do V. N., Lebrun M., Courtois B., & Gantet P. (2019). "Genome-wide Association Study of a Panel of Vietnamese Rice Landraces Reveals New QTLs for Tolerance to Water Deficit During the Vegetative Phase", *Rice*, 12(1), 4.
49. Hu C.-Y., Tsai Y.-Z., & Lin S.-F. (2014), "Development of STS and CAPS markers for variety identification and genetic diversity analysis of tea germplasm in Taiwan", *Botanical Studies*, 55.
50. Huang X., Qian Q., Liu Z., Sun H., He S., Luo D., Xia G., Chu C., Li J., & Fu X. (2009), "Natural variation at the DEP1 locus enhances grain yield in rice", *Nature Genetics*, 41, 494–497.
51. Huang X, Wei X, & Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B. (2010), "Genome-wide association studies of 14 agronomic traits in rice landraces", *Nature Genetics*, 42(11), 961–967.

52. Ikeda M., Miura K., Aya K., Kitano H., & Matsuoka M. (2013), "Genes offering the potential for designing yield-related traits in rice", *Current Opinion in Plant Biology*, 16(2), 213–220.
53. Ince A. G., Karaca M., & Elmasulu S. Y. (2014), "New microsatellite and CAPS-microsatellite markers for clarifying taxonomic and phylogenetic relationships within *Origanum L*", *Mol Breeding*, 13.
54. Jehan T., & Lakhanpaul S. (2006), "Single nucleotide polymorphism (SNP)–Methods and applications in plant genetics: A review", *INDIAN J BIOTECHNOL*, 26.
55. Jensen S. M., Svensgaard J., & Ritz C. (2020), "Estimation of the harvest index and the relative water content - Two examples of composite variables in agronomy", *European Journal of Agronomy*, 112, 125962.
56. Jia B., Zhao X., Qin Y., Irfan M., Kim T., Wang B., Wang S., & Keun Sohn J. (2019), "Quantitative trait loci mapping of panicle traits in rice", *Molecular Biology Research Communications*, 8(1), 9–15.
57. Kato T., Segami S., Toriyama M., Kono I., Ando T., Yano M., Kitano H., Miura K., & Iwasaki Y. (2011), "Detection of QTLs for grain length from large grain rice (*Oryza sativa L.*)", *Breeding Science*, 61(3), 269–274.
58. Kawahara Y., de la Bastide M., Hamilton J. P., Kanamori H., McCombie W. R., Ouyang S., Schwartz D. C., Tanaka T., Wu J., Zhou S., Childs K. L., Davidson R. M., Lin H., Quesada-Ocampo L., Vaillancourt B., Sakai H., Lee S. S., Kim J., Numa H., ... Matsumoto T. (2013), "Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data", *Rice*, 6(1):4.
59. Khlestkina E. K. & Salina E. A. (2006), *SNP markers: Methods of analysis, ways of development, and comparison on an example of common wheat*. 42, no. 6, pp. 725–736.

60. Kulkarni S. R., Balachandran S. M., Ulaganathan K., Balakrishnan D., Praveen M., Prasad A. S. H., Fiyaz R. A., Senguttuvel P., Sinha P., Kale R. R., Rekha G., Kousik M. B. V. N., Harika G., Anila M., Punniakoti E., Dilip T., Hajira S. K., Pranathi K., Das M. A., ... Sundaram R. M. (2020), "Molecular mapping of QTLs for yield related traits in recombinant inbred line (RIL) population derived from the popular rice hybrid KRH-2 and their validation through SNP genotyping", *Scientific Reports*, *10*(1), 13695.
61. Komatsu K., Maekawa M., Ujiie S., Satake Y., Furutani I., Okamoto H., Shimamoto K., & Kyojuka J. (2003), "LAX and SPA: Major regulators of shoot branching in rice", *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 11765–11770.
62. Konieczny A., & Ausubel F. M. (1993), "A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers", *The Plant Journal: For Cell and Molecular Biology*, *4*(2), 403–410.
63. Kyojuka J., Tokunaga H., & Yoshida A. (2014), "Control of grass inflorescence form by the fine-tuning of meristem phase change", *Current Opinion in Plant Biology*, *17*, 110–115.
64. Lekklar C., Pongpanich M., Suriya-Arunroj D., Chinpongpanich A., Tsai H., Comai L., Chadchawan S., & Buaboocha T. (2019), "Genome-wide association study for salinity tolerance at the flowering stage in a panel of rice accessions from Thailand", *BMC Genomics*, *20*(1), 76.
65. Li G., Cheng Y., Yin M., Yang J., Ying J., & Zhu C. (2021), "Detection of QTLs for panicle-related traits using an indica × japonica recombinant inbred line population in rice", *PeerJ*, *9*, e12504.
66. Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., & 1000 Genome Project Data Processing

- Subgroup. (2009), "The Sequence Alignment/Map format and SAMtools", *Bioinformatics (Oxford, England)*, 25(16), 2078–2079.
67. Li Y., Li X., Fu D., & Wu C. (2018), "Panicle Morphology Mutant 1 (PMM1) determines the inflorescence architecture of rice by controlling brassinosteroid biosynthesis", *BMC Plant Biology*, 18(1), 348.
68. Lim S.-H., & Ha S.-H. (2013), "Marker development for the identification of rice seed color", *Plant Biotechnology Reports*, 7(3), 391–398.
69. Lin X., Tang W., Ahmad S., Lu J., Colby C. C., Zhu J., & Yu Q. (2012), "Applications of targeted gene capture and next-generation sequencing technologies in studies of human deafness and other genetic disabilities", *Hearing Research*, 288(0).
70. Linh L.-H., Hang N.-T., Jin F.-X., Kang K.-H., Lee Y.-T., Kwon S.-J., & Ahn S.-N. (2008), "Introgression of a quantitative trait locus for spikelets per panicle from *Oryza minuta* to the *O. sativa* cultivar Hwaseongbyeon", *Plant Breeding*, 127(3), 262–267.
71. Liu T., Li L., Zhang Y., Xu C., Li X., & Xing Y. (2011), "Comparison of quantitative trait loci for rice yield, panicle length and spikelet density across three connected populations", *Journal of Genetics*, 90(2), 377–382.
72. Lu K.-T., Lee H.-C., Liu F.-S., Lo C.-F., & Lin J.-H. (2010), "Identification of Ginseng Radix in Chinese Medicine Preparations by Nested PCR-DNA Sequencing Method and Nested PCR-Restriction Fragment Length Polymorphism", *Journal of Food and Drug Analysis*, 18(1), 6.
73. Lu Y., Zhao G., Li Y., Fan J., Ding G., Zhao J., Ni X., Xu Y., & Wang W. (2013), "Identification of two novel waxy alleles and development of their molecular markers in sorghum", *Genome*, 56(5), 283–288.
74. Ma X., Feng F., Zhang Y., Elesawi I. E., Xu K., Li T., Mei H., Liu H., Gao N., Chen C., Luo L., & Yu S. (2019), "A novel rice grain size gene OsSNB

- was identified by genome-wide association study in natural population", *PLOS Genetics*, *15*(5), e1008191.
75. Mammadov J., Aggarwal R., Buyyarapu R., & Kumpatla S. (2012), "SNP markers and their impact on plant breeding", *International Journal of Plant Genomics*, *2012*, 728398.
 76. Marathi B., Guleria S., Mohapatra T., Parsad R., Mariappan N., Kurungara V. K., Atwal S. S., Prabhu K. V., Singh N. K., & Singh A. K. (2012), "QTL analysis of novel genomic regions associated with yield and yield related traits in new plant type based recombinant inbred lines of rice (*Oryza sativa*L.)", *BMC Plant Biology*, *12*(1), 137.
 77. Matsumoto T., Wu J., Itoh T., Numa H., Antonio B., & Sasaki T. (2016), "The Nipponbare genome and the next-generation of rice genomics research in Japan", *Rice*, *9*(1):33.
 78. McCouch S. R., Temnykh S., Lukashova A., Coburn J., DeClerck G., Cartinhour S., Harrington S., Thomson M., Septiningsih E., Semon M., Moncada P., & Li J. (2008), "Microsatellite markers in rice: Abundance, diversity, and applications". In *Rice Genetics IV: Vol. Volume 4*, pp. 117–135.
 79. Monat C., Tranchant-Dubreuil C., Kougbéadjó A., Farcy C., Ortega-Abboud E., Amanzougarene S., Ravel S., Agbessi M., Orjuela-Bouniol J., Summo M., & Sabot F. (2015), "TOGGLE: Toolbox for generic NGS analyses", *BMC Bioinformatics*, *16*, 374.
 80. Morandat F., Hill B., Osvald L., & Vitek J. (2012), "Evaluating the Design of the R Language. In J. Noble (Ed.)", *ECOOP 2012 – Object-Oriented Programming*, Vol. 7313, pp. 104–131.
 81. Mumtaz M., Saqib M., Abbas G., Akhtar J., & Qamar Z.-U.-Q. (2020), "Drought Stress Impairs Grain Yield and Quality of Different Rice

- Genotypes under Field Conditions by Impaired Photosynthetic Attributes and K Nutrition", *Rice Science*, 27, 5–9.
82. MyBaits-Hyb Capture Kits. (2020), Arbor Biosciences. Retrieved April 1, 2020.
 83. Neelam K., Brown-Guedira G., & Huang L. (2013), "Development and validation of a breeder-friendly KASPar marker for wheat leaf rust resistance locus Lr21", *Molecular Breeding*, 31(1), 233–237.
 84. Paradis E. (2010), "Pegas: An R package for population genetics with an integrated-modular approach", *Bioinformatics (Oxford, England)*, 26(3), 419–420.
 85. Peng S., Khush G. S., Virk P., Tang Q., & Zou Y. (2008), "Progress in ideotype breeding to increase rice yield potential", *Field Crops Research*, 108(1), 32–38.
 86. Peng S., & Khush G. (2003), "Four Decades of Breeding for Varietal Improvement of Irrigated Lowland Rice in the International Rice Research Institute", *Plant Production Science*, 6(3), 157–164.
 87. Peng Y., Gao Z., Zhang B., Liu C., Xu J., Ruan B., Hu J., Dong G., Guo L., Liang G., & Qian Q. (2014), "Fine mapping and candidate gene analysis of a major QTL for panicle structure in rice", *Plant Cell Reports*, 33(11), 1843–1850.
 88. Phung N. T. P., Mai C. D., Hoang G. T., Truong H. T. M., Lavarenne J., Gonin M., Nguyen K. L., Ha T. T., Do V. N., Gantet P., & Courtois B. (2016), "Genome-wide association mapping for root traits in a panel of rice accessions from Vietnam", *BMC Plant Biology*, 16(1), 64.
 89. Phung N. T. P., Mai C. D., Mournet P., Frouin J., Droc G., Ta N. K., Jouannic S., Lê L. T., Do V. N., Gantet P., & Courtois B. (2014), "Characterization of a panel of Vietnamese rice varieties using DArT and

- SNP markers for association mapping purposes", *BMC Plant Biology*, *14*(1), 371.
90. Qian Q., Guo L., Smith S. M., & Li J. (2016), "Breeding high-yield superior quality hybrid super rice by rational design", *National Science Review*, *3*(3), 283–294.
 91. Qian L., Hickey L. T., Stahl A., Werner C. R., Hayes B., Snowdon R. J., & Voss-Fels K. P. (2017), "Exploring and Harnessing Haplotype Diversity to Improve Yield Stability in Crops", *Frontiers in Plant Science*, *8*, 1534.
 92. R Core Team. (2019), *R: The R Project for Statistical Computing*. <https://www.r-project.org/>
 93. Ragavendran A., Vikas K. S., Vishnu V. N., Pallavi S., Ramchander S., Abhilash K. V., Arun K. S., Uma M., & Singh., Rajeev K V., Arvind K. (2019), "Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains", *Plant Biotechnology Journal*, *17*(8) DOI:10.1111/pbi.13087.
 94. Sakamoto T., & Matsuoka M. (2004), "Generating high-yielding varieties by genetic manipulation of plant architecture", *Current Opinion in Biotechnology*, *15*(2), 144–147.
 95. Sasaki K., Fujita D., Koide Y., Lumanglas P. D., Gannaban R. B., Tagle A. G., Obara M., Fukuta Y., Kobayashi N., & Ishimaru T. (2017), "Fine mapping of a quantitative trait locus for spikelet number per panicle in a new plant type rice and evaluation of a near-isogenic line for grain productivity", *Journal of Experimental Botany*, *68*(11), 2693–2702.
 96. Shapiro_test (2020): Shapiro-Wilk Normality Test in rstatix: Pipe-Friendly Framework for Basic Statistical Tests.

97. Si L., Chen J., Huang X., Gong H., Luo J., Hou Q., Zhou T., Lu T., Zhu J., Shanguan Y., Chen E., Gong C., Zhao Q., Jing Y., Zhao Y., Li Y., Cui L., Fan D., Lu Y., Han B. (2016), "OsSPL13 controls grain size in cultivated rice", *Nature Genetics*, 48(4), 447–456.
98. Sinha P., Singh V. K., Saxena R. K., Khan A. W., Abbai R., Chitikineni A., Desai A., Molla J., Upadhyaya H. D., Kumar A., & Varshney R. K. (2020), "Superior haplotypes for haplotype-based breeding for drought tolerance in pigeonpea (*Cajanus cajan* L.), *Plant Biotechnology Journal*, 18(12), 2482–2490.
99. Stram D. O. (2017), "Multi-SNP Haplotype Analysis Methods for Association Analysis. In R. C. Elston (Ed.)", *Statistical Human Genetics: Methods and Protocols*, pp. 485–504.
100. Sukumaran S., Dreisigacker S., Lopes M., Chavez P., & Reynolds M. P. (2015), "Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments", *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 128(2), 353–363.
101. Sun Z., Yin X., Ding J., Yu D., Hu M., Sun X., Tan Y., Sheng X., Liu L., Mo Y., Ouyang N., Jiang B., Yuan G., Duan M., Yuan D., & Fang J. (2017), "QTL analysis and dissection of panicle components in rice using advanced backcross populations derived from *Oryza Sativa* cultivars HR1128 and "Nipponbare", *PloS One*, 12(4), e0175692.
102. Ta K. N., Khong N. G., Ha T. L., Nguyen D. T., Mai D. C., Hoang T. G., Phung T. P. N., Bourrie I., Courtois B., Tran T. T. H., Dinh B. Y., La T. N., Do N. V., Lebrun M., Gantet P., & Jouannic S. (2018), "A genome-wide association study using a Vietnamese landrace panel of rice (*Oryza sativa*)

- reveals new QTLs controlling panicle morphological traits", *BMC Plant Biology*, 18(1), 282.
103. Tan Y.-Y., Fu H.-W., Zhao H.-J., Lu S., Fu J.-J., Li Y.-F., Cui H.-R., & Shu Q.-Y. (2013), "Functional molecular markers and high-resolution melting curve analysis of low phytic acid mutations for marker-assisted selection in rice", *Molecular Breeding*, 31(3), 517–528.
 104. Tanaka W., Toriba T., & Hirano H.-Y. (2014), "Chapter Eight-Flower Development in Rice. In F. Fornara (Ed.)", *Advances in Botanical Research*, Vol. 72, pp. 221–262, Academic Press.
 105. Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S., & McCouch S. (2001), "Computational and Experimental Analysis of Microsatellites in Rice (*Oryza sativa* L.): Frequency, Length Variation, Transposon Associations, and Genetic Marker Potential", *Genome Research*, 11(8), 1441–1452.
 106. Temnykh S., Park W. D., Ayres N., Cartinhour S., Hauck N., Lipovich L., Cho Y. G., Ishii T., & McCouch S. R. (2000), "Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.)", *Theoretical and Applied Genetics*, 100(5), 697–712.
 107. Thomson M. J., Tai T. H., McClung A. M., Lai X.-H., Hinga M. E., Lobos K. B., Xu Y., Martinez C. P., & McCouch S. R. (2003), "Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson", *Theoretical and Applied Genetics*, 107(3), 479–493.
 108. Toparslan E., Karabag K., & Bilge U. (2020), "A workflow with R: Phylogenetic analyses and visualizations using mitochondrial cytochrome b gene sequences", *PLOS ONE*, 15(12), e0243927.

109. Ui H., Sameri M., Pourkheirandish M., Chang M.-C., Shimada H., Stein N., Komatsuda T., & Handa H. (2015), "High-resolution genetic mapping and physical map construction for the fertility restorer Rfm1 locus in barley", *Theoretical and Applied Genetics*, 128(2), 283–290.
110. Van der Auwera, G. A., Carneiro M. O., Hartl C., Poplin R., del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella K. V., Altshuler D., Gabriel S., & DePristo M. A. (2013), "From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline", *Current Protocols in Bioinformatics*, 43, 11.10.1-11.10.33.
111. Wang S., Wong D., Forrest K., Allen A., Chao S., Huang B. E., Maccaferri M., Salvi S., Milner S. G., Cattivelli L., Mastrangelo A. M., Whan A., Stephen S., Barker G., Wieseke R., Plieske J., Lillemo M., Mather D., Appels R., Akhunov E. (2014), "Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array", *Plant Biotechnology Journal*, 12(6), 787–796.
112. Wang W., Xu X., Zhu C., Gu J., Zhang W., Liu G., & Zhu J. (2019), "Elevated CO₂-induced changes in cytokinin and nitrogen metabolism are associated with different responses in the panicle architecture of two contrasting rice genotypes", *Plant Growth Regulation*, 89(2), 119–129.
113. Weng J., Gu S., Wan X., Gao H., Guo T., Su N., Lei C., Zhang X., Cheng Z., Guo X., Wang J., Jiang L., Zhai H., & Wan J. (2008), "Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight", *Cell Research*, 18, 1199–1209.
114. Wickham H., Hester J., Chang W., Bryan J., & RStudio. (2021), *devtools: Tools to Make Developing R Packages Easier* (2.4.3).

115. Wu Y., Huang M., Tao X., Guo T., Chen Z., & Xiao W. (2016), "Quantitative trait loci identification and meta-analysis for rice panicle-related traits", *Molecular Genetics and Genomics: MGG*, 291(5), 1927–1940.
116. Xing Y., & Zhang Q. (2010), "Genetic and molecular bases of rice yield", *Annual Review of Plant Biology*, 61, 421–442.
117. Xu X., Zhang M., Xu Q., Feng Y., Yuan X., Yu H., Wang Y., Wei X., & Yang Y. (2020), "Quantitative trait loci identification and genetic diversity analysis of panicle structure and grain shape in rice", *Plant Growth Regulation*, 90(1), 89–100.
118. Xue W., Xing Y., Weng X., Zhao Y., Tang W., Wang L., Zhou H., Yu S., Xu C., Li X., & Zhang Q. (2008), "Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice", *Nature Genetics*, 40(6), 761–767.
119. Ya-fang Z., Yu-yin M., Zong-xiang C., Jie Z., Tian-xiao C., Qian-qian L., Xue-biao P., & Shi-min Z. (2015), "Genome-Wide Association Studies Reveal New Genetic Targets for Five Panicle Traits of International Rice Varieties", *Rice Science*, 22(5), 217–226.
120. Yoshida H., & Nagato Y. (2011), "Flower development in rice", *Journal of Experimental Botany*, 62(14), 4719–4730.
121. Yu H., Qiu Z., Xu Q., Wang Z., Zeng D., Hu J., Zhang G., Zhu L., Gao Z., Chen G., Guo L., Qian Q., & Ren D. (2017), "Fine mapping of LOW TILLER 1, a gene controlling tillering and panicle branching in rice", *Plant Growth Regulation*, 83(1), 93–104.
122. Yuan L. (2015), "Development of Hybrid Rice for Worldwide Food Security", *Engineering*, 5.
123. Zeng Z.-B. (2001), "QTL Mapping", In *Brenner's Encyclopedia of Genetics* (pp. 8–12). Elsevier.

124. Zhang J., Chiodini R., Badr A., & Zhang G. (2011), "The impact of next-generation sequencing on genomics", *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, 38(3), 95–109.
125. Zhang Y.S., Luo L.J., Xu C.G., Zhang Q.F., Xing Y.Z. (2006), "Quantitative trait loci for panicle size, heading date and plant height cosegregating in trait-performance derived near-isogenic lines of rice (*Oryza sativa*)", *Theoretical and Applied Genetics*, 113, 361–368
126. Zhao C. F., Chen T., Zhao Q. Y., Zhou L. H., Zhao L., Zhang Y. D., Zhu Z., Yao S., & Wang C. L. (2016), "Analysis of QTLs for panicle exertion and its relationship with yield and yield-related traits in rice (*Oryza sativa* L.)", *Genetics and Molecular Research: GMR*, 15(2).
127. Zhao K., Tung C.-W., Eizenga G. C., Wright M. H., Ali M. L., Price A. H., Norton G. J., Islam M. R., Reynolds A., Mezey J., McClung A. M., Bustamante C. D., & McCouch S. R. (2011), "Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*", *Nature Communications*, 2, 467.
128. Zhu Z., Li X., Wei Y., Guo S., & Sha A. (2018), "Identification of a Novel QTL for Panicle Length From Wild Rice (*Oryza minuta*) by Specific Locus Amplified Fragment Sequencing and High Density Genetic Mapping", *Frontiers in Plant Science*, 9, 1492.
129. Zhu M., Liu D., Liu W., Li D., Liao Y., Li J., Fu C., Fu F., Huang H., Zeng X., Ma X., & Wang F. (2017), "QTL mapping using an ultra-high-density SNP map reveals a major locus for grain yield in an elite rice restorer R998", *Scientific Reports*, 7(1), 10914.
130. Zhuang JY, Fan YY, Wu JL, Xia YW, & Zheng KL. (2001), "Comparison of the detection of QTL for yield traits in different generations of a rice

cross using two mapping approaches", Europe PMC. *PMID: 11441659*, 28(5):458-464.

C. Tài liệu web

131. <http://www.gramene.org>
132. <http://fao.org>
133. <http://irri.org>
134. <http://passel.unl.edu/pages/>
135. <https://vi.wikipedia.org/w/index.php?title=L%C3%B4-cut> (Lô-cut tính trạng số lượng)
136. <http://rice.plantbiology.msu.edu>
137. <http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>
138. <http://ncbi.nlm.nih.gov>
139. Tổng cục thống kê (2022), <https://www.gso.gov.vn/du-lieu-va-so-lieu-thong-ke/2022/>

PHỤ LỤC 1
DANH SÁCH VÀ TRÌNH TỰ CÁC HAPLOTYPE TRONG VÙNG QTL9
CỦA TẬP ĐOÀN 155 GIỐNG LÚA BẢN ĐỊA VIỆT NAM

H	STT giống	Mã giống	Nhóm	Thời gian sinh trưởng		Số hạt/ bông		Số gié thứ cấp/ bông		Trình tự haplotype
				Năm	Năm	Năm	Năm	Năm	Năm	
				2014	2015	2014	2015	2014	2015	
H9	1	ACZ	J	81,5	79	185,50	217,33	30,89	37,67	atataagaa
H1	2	G1	I	73,5	69,5	124,89	186,22	24,72	39,00	gagagcgaa
	3	G10	I	97	109,5	144,39	169,67	22,50	24,00	gagagcgaa
	4	G100	J	96,5	103	207,22	244,28	39,28	47,28	gagagcgaa
	6	G102	I	113,5	90	160,80	249,50	28,47	48,17	gagagcgaa
	8	G104	I	100,5	111	201,33	193,56	35,17	30,61	gagagcgaa
	10	G106	J	108,5	115	165,67	206,72	31,39	40,17	gagagcgaa
	11	G107	J	111	113,5	114,94	128,72	20,11	25,78	gagagcgaa
	13	G11	I	101,5	109,5	176,61	178,06	31,67	32,06	gagagcgaa
	16	G113	I	96	110,5	163,39	186,11	29,50	34,83	gagagcgaa
	17	G115	I	79	74	246,28	236,39	50,28	45,56	gagagcgaa
	18	G117	J	109,5	113	125,00	184,44	20,61	33,89	gagagcgaa
	19	G119	I	109,5	121,5	127,65	136,33	22,12	21,89	gagagcgaa
	20	G12	I	107,5	107	117,83	131,28	16,28	19,06	gagagcgaa
	21	G124	J	107	113,5	140,28	190,17	25,61	36,61	gagagcgaa
	22	G125	I	80,5	77	173,17	163,44	31,56	24,78	gagagcgaa
	23	G126	J	111	116,5	205,06	242,33	35,78	37,11	gagagcgaa
	24	G128	J	107,5	112,5	164,50	226,72	27,50	38,28	gagagcgaa
	25	G129	I	84	82,5	271,50	293,00	56,22	53,56	gagagcgaa
	29	G133	I	75,5	69,5	167,33	163,22	38,78	35,00	gagagcgaa
	31	G138	I	151	153,5	203,50	225,83	38,06	40,06	gagagcgaa
	32	G14	I	102,5	92,5	115,33	141,56	15,94	19,28	gagagcgaa
	33	G140	I	150	155	254,89	289,28	49,78	54,67	gagagcgaa
	34	G143	I	150,5	157	255,94	249,22	49,17	47,22	gagagcgaa
	35	G146	I	140	145	179,89	181,00	33,89	34,00	gagagcgaa
	36	G150	I	155,5	156,5	130,22	207,72	21,11	39,50	gagagcgaa
	37	G152	J	85	85	202,72	292,11	39,17	56,11	gagagcgaa

H	STT giống	Mã giống	Nhóm	Thời gian sinh trưởng		Số hạt/ bông		Số gié thứ cấp/ bông		Trình tự haplotype
				Năm	Năm	Năm	Năm	Năm	Năm	
				2014	2015	2014	2015	2014	2015	
39	G154	J	84,5	81	133,72	161,17	28,39	34,11	gagagcgaa	
44	G16	J	104,5	111	160,59	213,78	31,82	37,28	gagagcgaa	
47	G165	I	74	73	146,28	147,00	30,72	29,00	gagagcgaa	
48	G167	I	162	162	174,00	178,83	32,33	35,28	gagagcgaa	
49	G168	I	72,5	69,5	157,22	134,83	31,72	24,56	gagagcgaa	
51	G17	I	107	109	134,78	156,17	20,67	24,50	gagagcgaa	
52	G170	I	76	67,5	159,44	137,50	30,72	25,06	gagagcgaa	
53	G171	I	75,5	72	157,50	156,28	35,11	32,33	gagagcgaa	
55	G173	I	79	72,5	186,28	238,11	38,67	48,72	gagagcgaa	
56	G177	J	98	105	163,89	215,67	30,06	43,06	gagagcgaa	
57	G178	J	97,5	109	179,00	220,72	36,06	45,83	gagagcgaa	
59	G18	I	108,5	109,5	101,00	123,56	12,67	17,17	gagagcgaa	
64	G187	J	85	78	137,11	195,67	26,17	37,22	gagagcgaa	
66	G19	I	95,5	100,5	133,94	176,83	24,22	31,11	gagagcgaa	
68	G191	J	104	113	176,61	208,78	34,56	39,94	gagagcgaa	
69	G192	I	109	108	146,73	180,56	21,13	32,44	gagagcgaa	
70	G193	J	94	91,5	152,94	213,00	29,06	44,78	gagagcgaa	
71	G194	J	85	85,5	171,22	219,06	34,56	46,56	gagagcgaa	
72	G195	J	91	97	182,06	320,50	34,94	62,17	gagagcgaa	
73	G2	I	83,5	84	227,83	249,44	46,06	49,56	gagagcgaa	
75	G200	J	84,5	80	215,83	233,11	44,89	43,61	gagagcgaa	
76	G201	I	93	92,5	290,44	328,67	53,67	58,33	gagagcgaa	
77	G202	J	99,5	111,5	208,39	358,22	37,72	63,33	gagagcgaa	
78	G203	J	90,5	89,5	167,89	223,56	34,06	50,17	gagagcgaa	
79	G204	J	92,5	97,5	196,89	329,50	36,83	61,56	gagagcgaa	
81	G206	J	104,5	107,5	201,83	196,22	38,22	37,50	gagagcgaa	
82	G207	I	74	67	157,39	138,22	33,61	27,17	gagagcgaa	
85	G21	I	107,5	108,5	140,94	147,61	23,44	22,72	gagagcgaa	
86	G210	J	95,5	98,5	194,39	210,28	41,33	42,67	gagagcgaa	
87	G212	J	101,5	103,5	167,94	235,33	29,67	44,50	gagagcgaa	

H	STT giống	Mã giống	Nhóm	Thời gian sinh trưởng		Số hạt/ bông		Số gié thứ cấp/ bông		Trình tự haplotype
				Năm	Năm	Năm	Năm	Năm	Năm	
				2014	2015	2014	2015	2014	2015	
88	G214	J	90	89,5	156,11	160,22	30,50	31,28	gagagcgaa	
89	G216	J	84,5	87,5	204,72	279,94	44,44	59,83	gagagcgaa	
90	G217	J	83	80	123,17	170,00	24,11	32,44	gagagcgaa	
91	G219	I	111	115,5	131,44	191,00	19,17	30,06	gagagcgaa	
92	G22	I	109,5	113,5	139,72	179,06	24,67	32,44	gagagcgaa	
93	G220	J	100,5	111,5	201,00	261,72	36,33	46,17	gagagcgaa	
94	G221	J	97	100	169,50	230,78	32,94	47,11	gagagcgaa	
95	G222	J	99	103,5	159,00	175,83	29,67	34,39	gagagcgaa	
96	G223	J	97	91,5	135,00	223,06	26,06	41,94	gagagcgaa	
97	G25	J	99	107	189,94	209,22	35,61	40,28	gagagcgaa	
98	G26	J	105	108,5	177,11	211,33	33,11	41,11	gagagcgaa	
99	G299	J	105,5	110	176,33	248,39	34,06	47,67	gagagcgaa	
100	G3	I	92,5	99	140,56	161,39	23,00	23,61	gagagcgaa	
102	G38	J	100,5	112,5	185,17	224,83	35,39	43,17	gagagcgaa	
104	G4	I	94,5	102,5	121,72	162,22	20,50	26,89	gagagcgaa	
105	G45	J	73	74,5	148,33	169,28	30,67	34,78	gagagcgaa	
106	G46	J	80	80	141,50	171,17	28,83	33,83	gagagcgaa	
107	G47	J	81,5	75,5	142,33	163,17	24,94	27,11	gagagcgaa	
108	G48	J	82,5	82	158,67	235,39	31,67	46,17	gagagcgaa	
110	G5	I	131,5	133	171,72	196,33	31,00	34,44	gagagcgaa	
111	G50	J	94	100	175,28	206,28	33,72	41,83	gagagcgaa	
112	G51	I	89,5	96	178,65	225,56	35,59	42,83	gagagcgaa	
113	G52	I	89,5	91,5	147,50	177,61	28,17	32,94	gagagcgaa	
114	G53	I	132,5	134,5	216,33	317,44	36,94	57,22	gagagcgaa	
117	G57	I	122	131	204,67	193,67	40,72	38,33	gagagcgaa	
118	G58	I	100	108	186,78	185,78	31,94	29,28	gagagcgaa	
119	G59	I	120,5	129	179,56	172,11	34,83	31,50	gagagcgaa	
120	G6	I	96,5	101	130,22	187,33	23,67	34,83	gagagcgaa	
121	G61	J	82	76	144,63	165,33	30,25	34,72	gagagcgaa	
122	G62	I	97,5	103,5	176,83	211,67	37,00	42,89	gagagcgaa	

H	STT giống	Mã giống	Nhóm	Thời gian sinh trưởng		Số hạt/ bông		Số gié thứ cấp/ bông		Trình tự haplotype
				Năm 2014	Năm 2015	Năm 2014	Năm 2015	Năm 2014	Năm 2015	
				127	G69	I	130,5	135,5	204,94	
128	G7	I	100	103	127,39	155,56	22,67	27,44	gagagcgaa	
129	G70	I	131	139	228,11	309,94	44,28	60,39	gagagcgaa	
130	G72	I	122,5	133	198,06	174,11	35,67	29,89	gagagcgaa	
131	G73	I	117	121,5	162,83	188,17	27,44	30,78	gagagcgaa	
132	G74	I	126,5	134	183,28	227,89	30,22	40,33	gagagcgaa	
133	G77	I	127,5	132	199,78	201,22	36,94	35,50	gagagcgaa	
135	G79	I	129,5	135,5	284,00	239,11	51,61	39,06	gagagcgaa	
136	G8	I	97,5	101	137,61	181,39	25,61	32,61	gagagcgaa	
140	G85	J	77,5	74	117,17	143,06	21,00	26,56	gagagcgaa	
141	G86	J	100	107,5	186,83	203,78	37,11	41,00	gagagcgaa	
142	G87	J	84	82	145,11	203,11	26,89	37,44	gagagcgaa	
143	G88	J	89	97	208,17	248,89	40,11	46,61	gagagcgaa	
144	G89	J	91	96	127,39	204,22	24,22	41,56	gagagcgaa	
145	G9	I	93	100	131,61	166,00	23,78	28,33	gagagcgaa	
146	G90	J	105,5	107	171,61	219,67	32,39	43,17	gagagcgaa	
147	G91	J	103	110,5	160,39	244,18	28,17	44,53	gagagcgaa	
148	G92	J	108	105	171,50	246,44	30,67	46,61	gagagcgaa	
149	G93	I	90,5	96	171,72	221,72	34,11	45,33	gagagcgaa	
150	G94	I	92	99,5	186,64	229,83	35,93	45,56	gagagcgaa	
151	G95	I	104,5	103,5	124,11	173,44	20,67	32,00	gagagcgaa	
152	G96	I	86,5	79,5	139,11	149,44	23,33	25,72	gagagcgaa	
153	G98	J	76	72	173,06	230,94	33,94	45,94	gagagcgaa	
154	G99	I	103,5	107	161,78	194,78	31,00	37,83	gagagcgaa	
155	IR64	I	79	77,5	143,50	144,67	31,17	29,28	gagagcgaa	
H2	5	G101	J	114,5	118	252,00	266,56	54,59	55,67	atataaatt
	9	G105	I	100	110,5	210,56	233,00	36,56	41,22	atataaatt
	14	G110	I	123	131	193,44	224,44	35,83	41,11	atataaatt
	15	G111	I	134,5	110,5	227,00	256,11	43,39	47,78	atataaatt
	26	G130	J	136,5	141	216,33	295,06	36,28	55,67	atataaatt

H	STT giống	Mã giống	Nhóm	Thời gian sinh trưởng		Số hạt/ bông		Số gié thứ cấp/ bông		Trình tự haplotype
				Năm 2014	Năm 2015	Năm 2014	Năm 2015	Năm 2014	Năm 2015	
	27	G131	J	135,5	141	210,50	287,78	36,17	56,00	atataaatt
	28	G132	I	134,5	140,5	264,89	293,94	57,94	62,67	atataaatt
	30	G134	J	132	132	142,89	217,28	17,39	38,44	atataaatt
	38	G153	I	94	103	206,88	227,17	39,41	45,06	atataaatt
	40	G155	I	94,5	102	227,44	244,33	43,17	47,94	atataaatt
	41	G156	I	116	122	293,65	276,94	55,35	54,28	atataaatt
	43	G158	J	136	141	237,78	290,06	39,78	52,72	atataaatt
	45	G160	I	83	75	218,71	159,94	44,29	32,39	atataaatt
	46	G161	I	135,5	142	232,78	263,83	45,56	52,56	atataaatt
	50	G169	I	71,5	72,5	209,06	177,89	42,83	34,17	atataaatt
	60	G181	I	101	109	212,39	207,44	43,22	42,72	atataaatt
	61	G182	I	81,5	81	206,50	277,44	48,17	57,33	atataaatt
	62	G183	I	82	78	185,67	230,11	35,22	43,17	atataaatt
	63	G186	I	88	84	239,78	248,22	51,78	50,61	atataaatt
	65	G189	I	94,5	79	250,50	228,22	48,67	45,22	atataaatt
	67	G190	I	90,5	86,5	214,39	240,72	42,22	45,94	atataaatt
	80	G205	I	99	108	220,33	244,22	44,44	47,39	atataaatt
	83	G208	I	111	116	203,28	223,50	36,83	38,50	atataaatt
	84	G209	I	89	84,5	229,83	300,22	50,67	60,44	atataaatt
	123	G63	I	79	82	168,28	221,33	30,11	40,11	atataaatt
	126	G67	I	158	162	234,50	303,72	48,44	59,22	atataaatt
	137	G80	J	135,5	140	211,78	288,67	37,44	52,78	atataaatt
	138	G83	J	125,5	134	214,06	224,67	36,50	39,61	atataaatt
	139	G84	J	135,5	140	247,11	328,94	44,44	61,78	atataaatt
H3	7	G103	J	121	119	143,39	235,06	24,72	44,61	atataaata
	42	G157	J	93	87	186,39	243,00	34,94	48,61	atataaata
H4	12	G109	I	113	121,5	191,44	199,11	30,22	32,89	gagagcggtt
	115	G54	I	102,5	110,5	216,22	213,33	36,56	34,72	gagagcggtt
	134	G78	I	130	130	177,83	259,33	32,33	48,78	gagagcggtt
H5	54	G172	I	75	68,5	204,44	192,67	43,56	41,83	gtgagcgaa

H	STT giống	Mã giống	Nhóm	Thời gian sinh trưởng		Số hạt/ bông		Số gié thứ cấp/ bông		Trình tự haplotype
				Năm 2014	Năm 2015	Năm 2014	Năm 2015	Năm 2014	Năm 2015	
	109	G49	I	88	83	177,11	254,89	27,50	43,39	gtgagcgaa
H6	58	G179	J	93	98,5	205,28	284,39	38,89	55,17	gagtgcgaa
	74	G20	I	105	106	149,61	134,00	22,61	19,22	gagtgcgaa
H7	101	G37	I	116,5	122	236,17	207,94	46,61	40,67	gtgagcatt
	103	G39	I	98	75,5	119,78	133,33	22,83	26,33	gtgagcatt
H8	116	G56	I	94,5	98	183,61	210,11	37,06	46,17	gagagcgat
	124	G64	I	91,5	91	209,56	256,56	38,83	48,33	gagagcgat
	125	G65	I	94	105,5	175,61	239,06	32,06	45,78	gagagcgat

Ghi chú: H: Haplotye, STT: số thứ tự, I: Indica, J: Japonica.

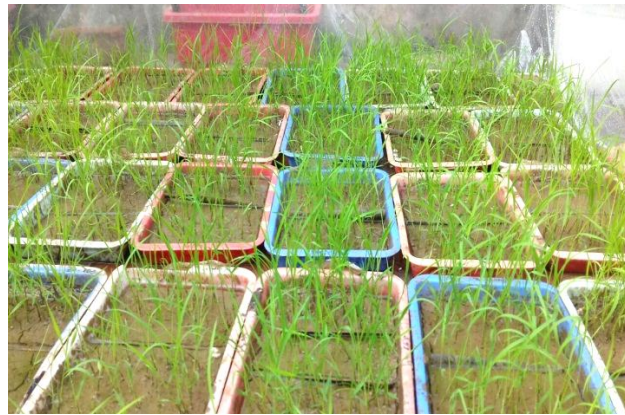
PHỤ LỤC 2
MỘT SỐ HÌNH ẢNH THỰC HIỆN ĐỀ TÀI LUẬN ÁN



Quần thể F_2 giai đoạn mạ (tháng 02/2018)



Các dòng F_2 trồng tại nhà lưới Viện Di truyền Nông nghiệp (tháng 5/2018)



Quần thể F_3 giai đoạn mạ (tháng 01/2019)



Bố trí thí nghiệm đồng ruộng (tháng 02/2019)



Khu thí nghiệm đánh giá quần thể F_3 ngoài đồng ruộng (tại Đông La, Hoài Đức, Hà Nội, vụ xuân năm 2019)





Quần thể F₃ ngoài đồng ruộng sau cấy 01 tháng (tháng 3/2019)



Quần thể F₃ ngoài đồng ruộng sau cấy 02 tháng (tháng 4/2019)



Hoa lúa



Quần thể F₃ giai đoạn nở hoa (tháng 4-5/2019)



Đánh giá và thu bông quần thể F₃ ngoài đồng ruộng (tháng 6/2019)

PHỤ LỤC 3: KẾT QUẢ ĐIỆN DI MAO QUẢN QIAXCEL GEN ĐỂ XÁC ĐỊNH CÁC DÒNG F2 ĐỒNG HỢP TỬ

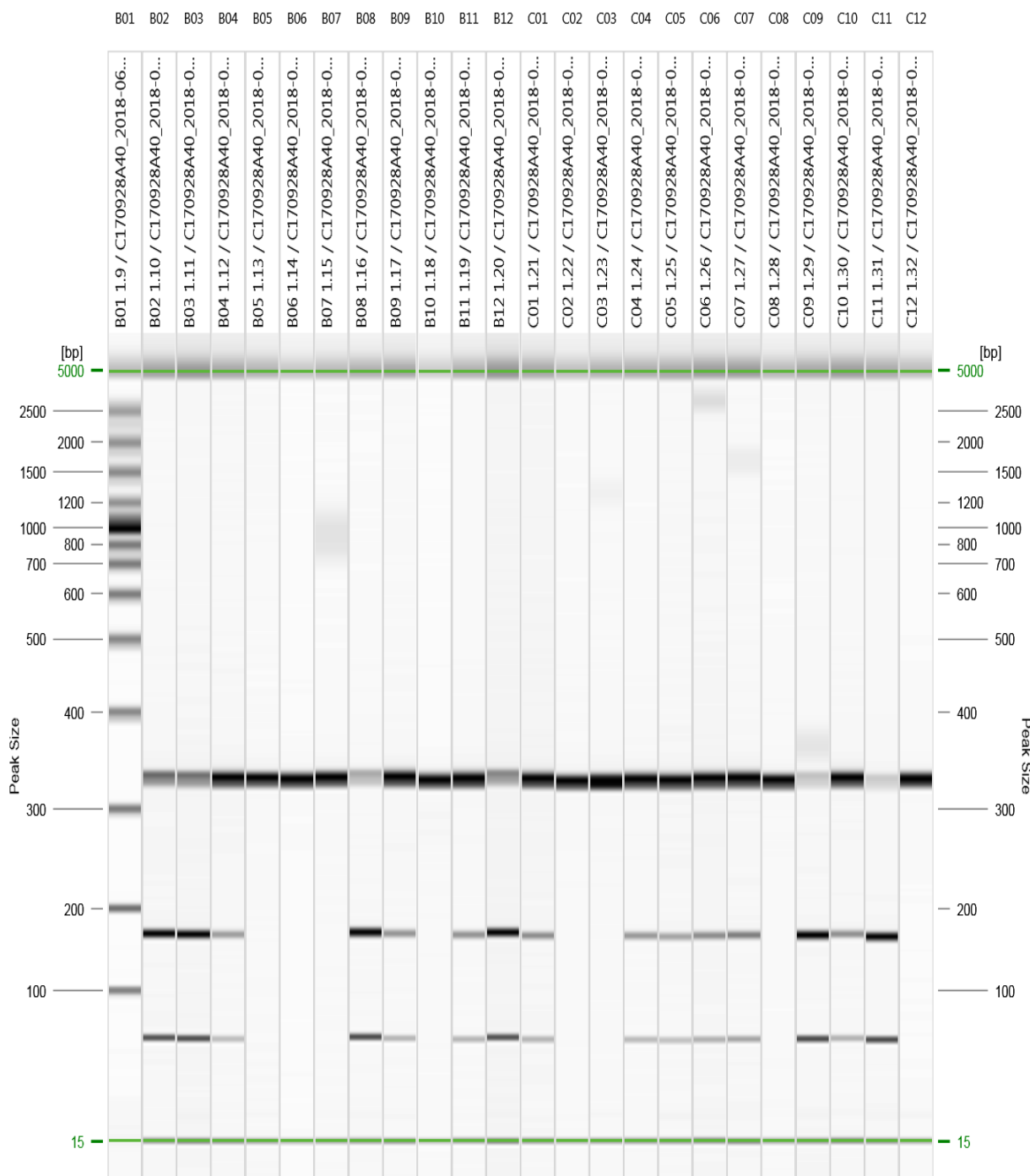


Figure: 1



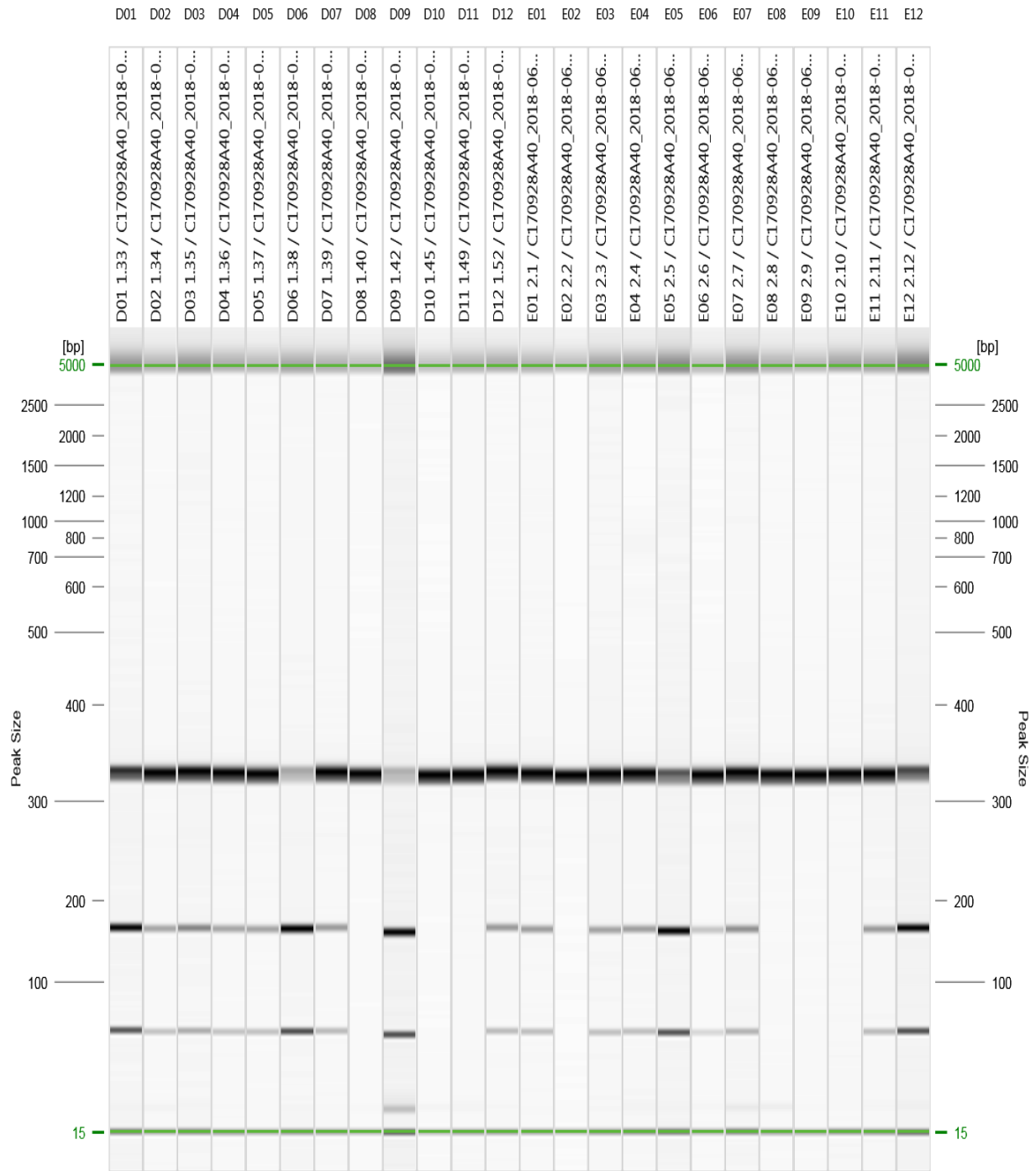


Figure: 2

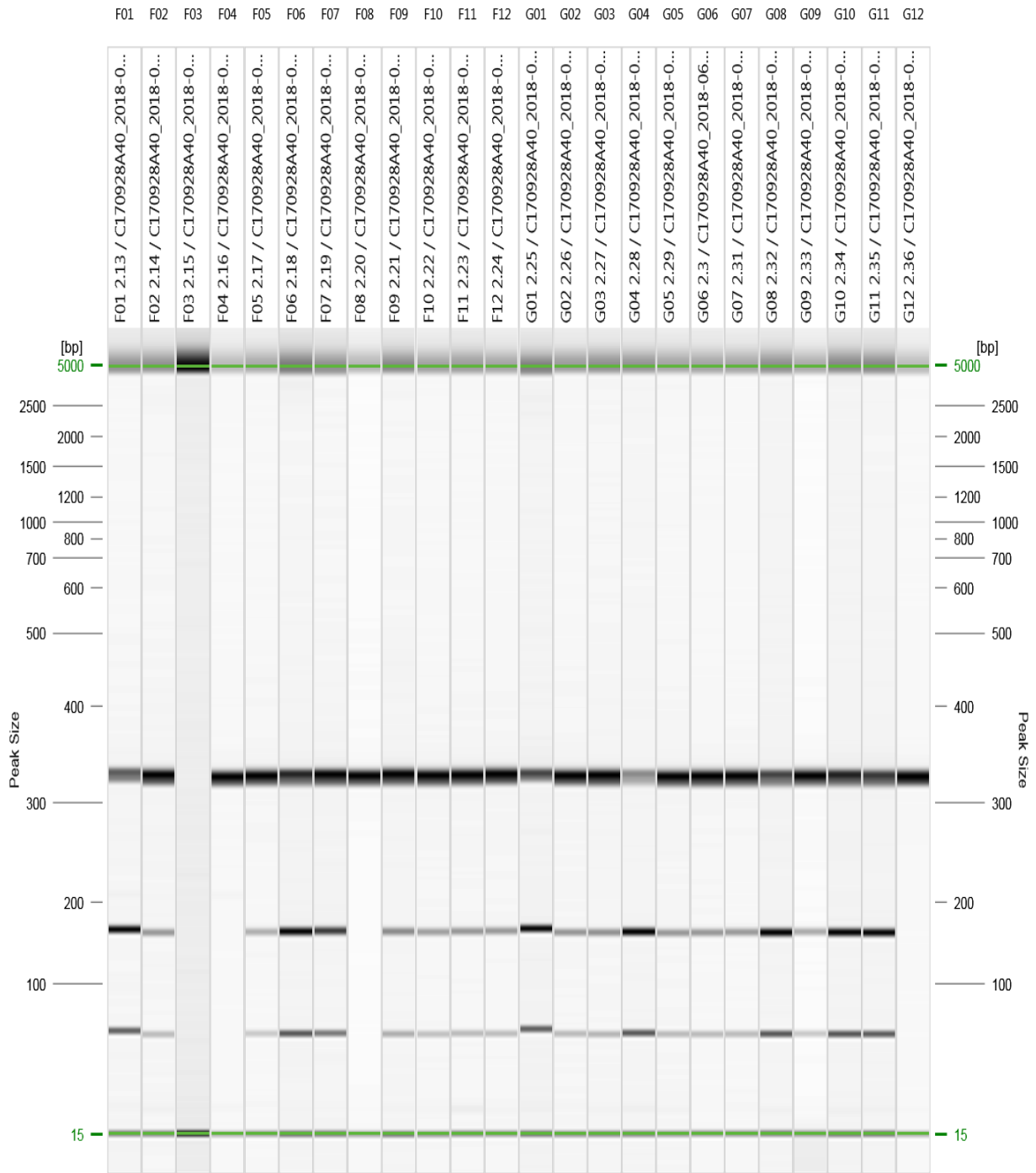


Figure: 3

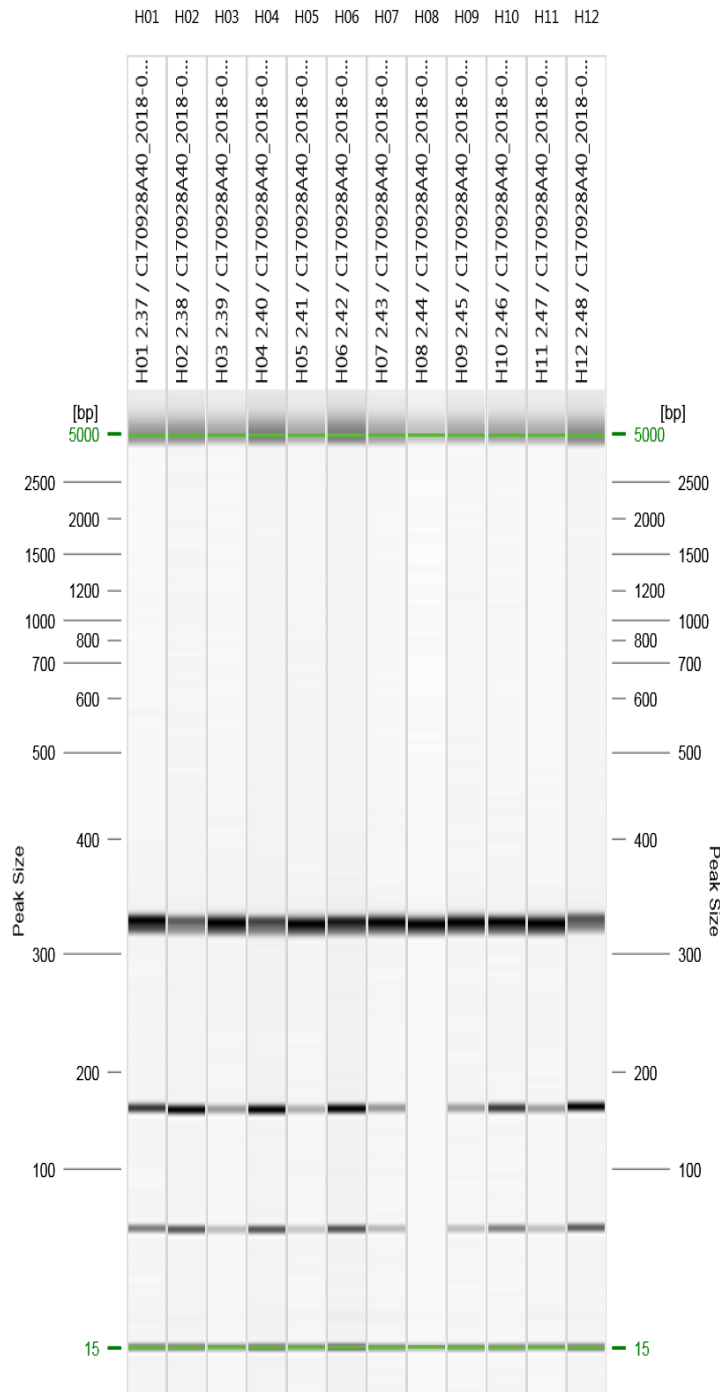


Figure: 4

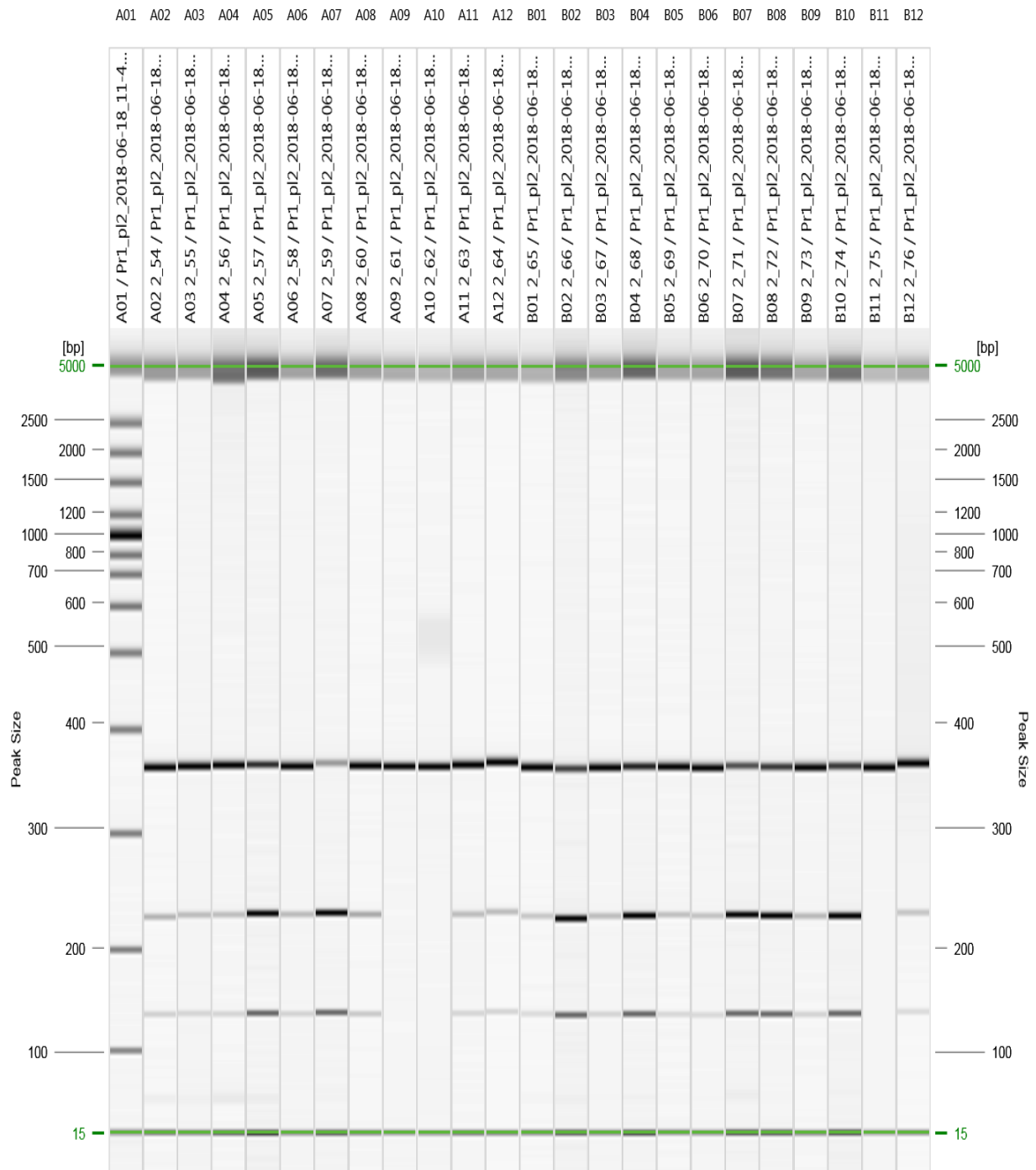


Figure: 1

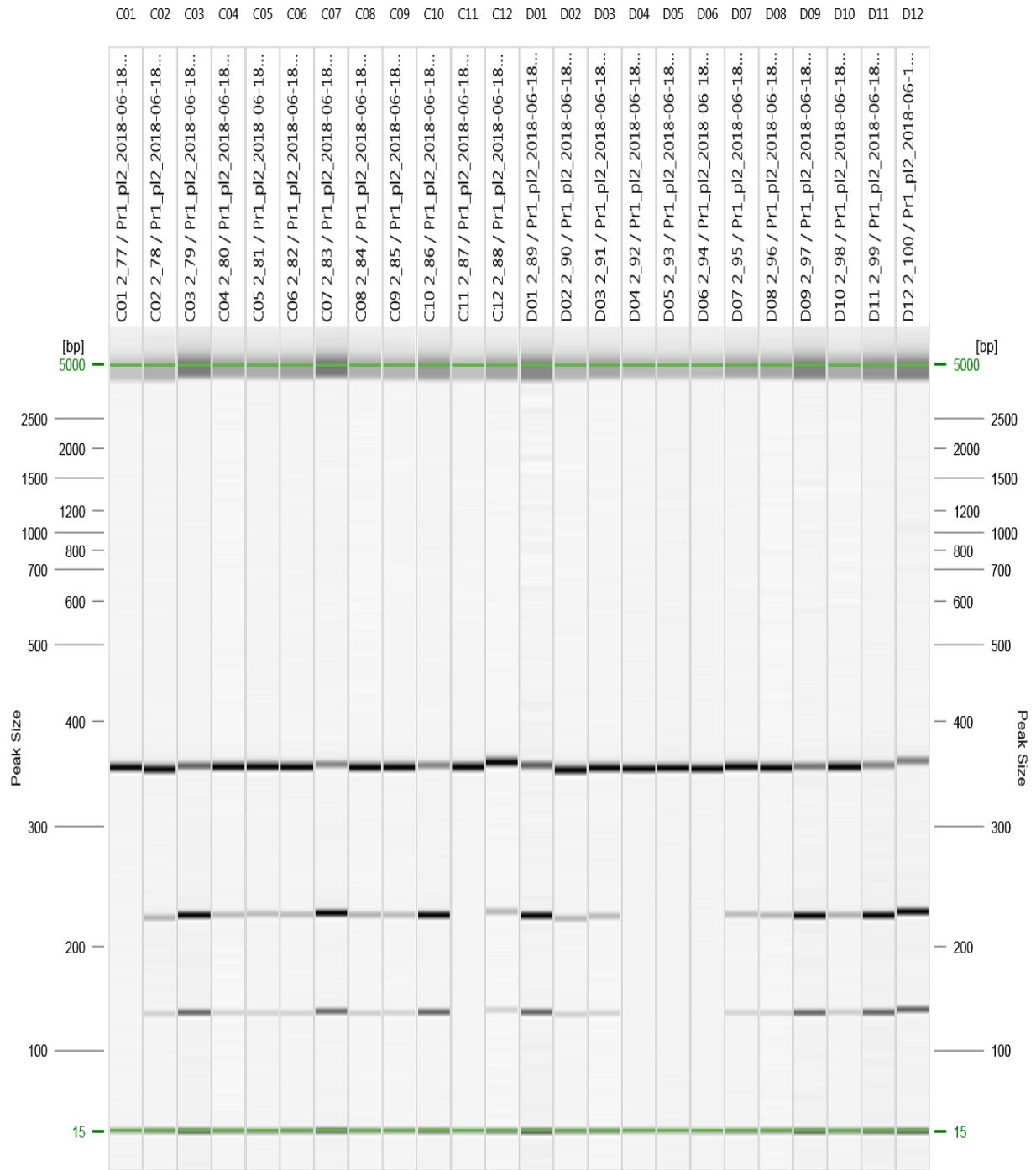


Figure: 2

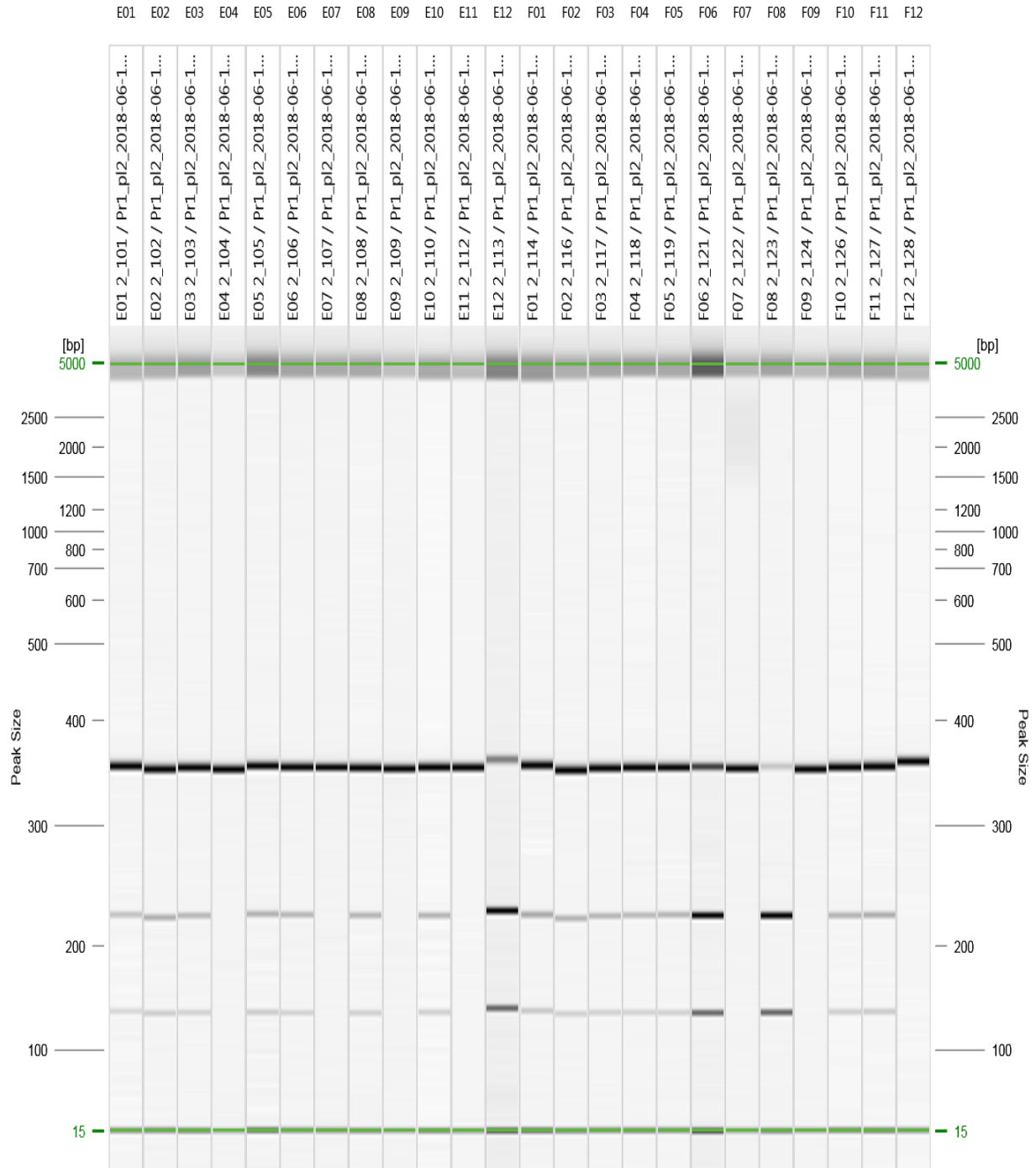


Figure: 3

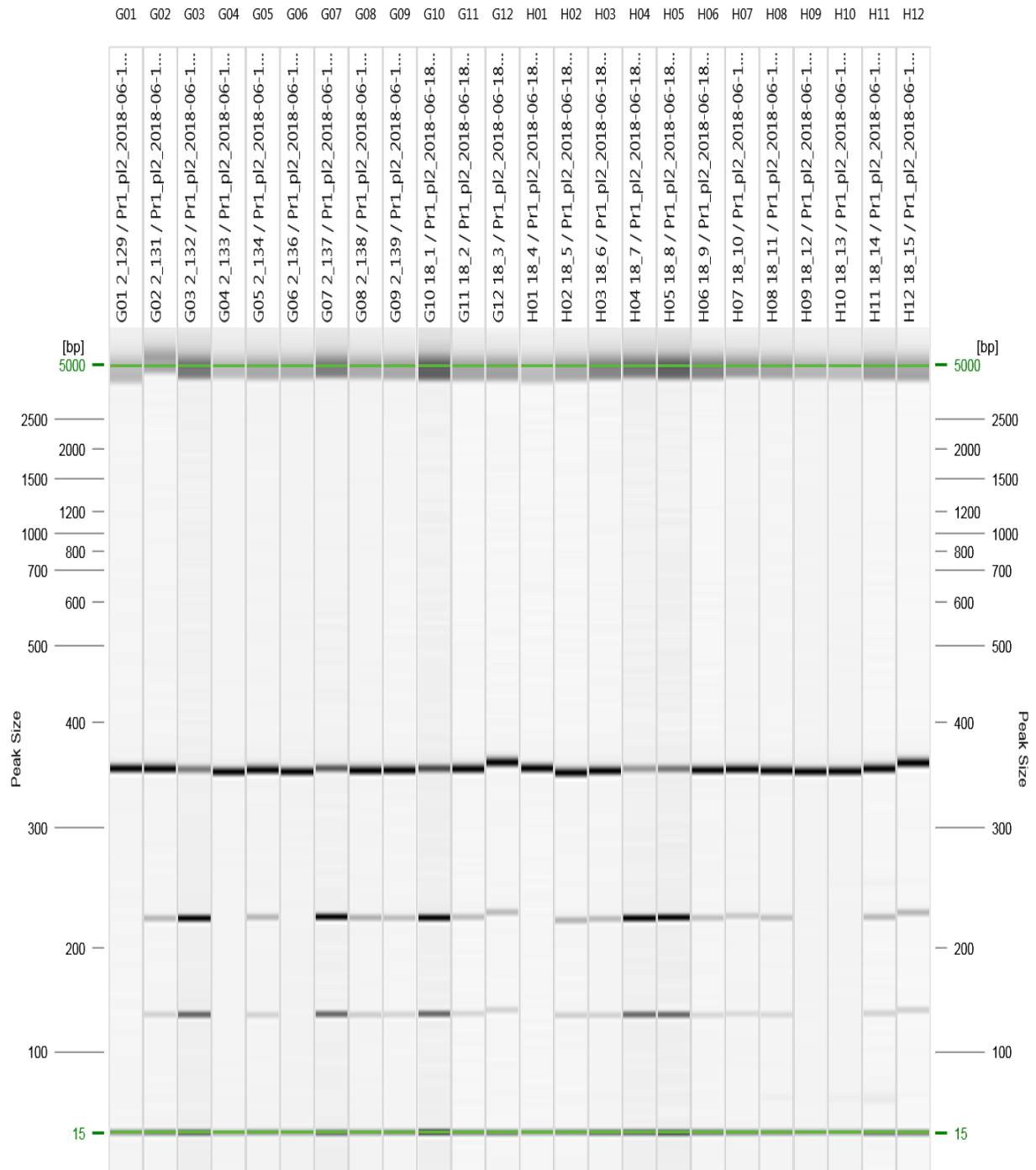


Figure: 4

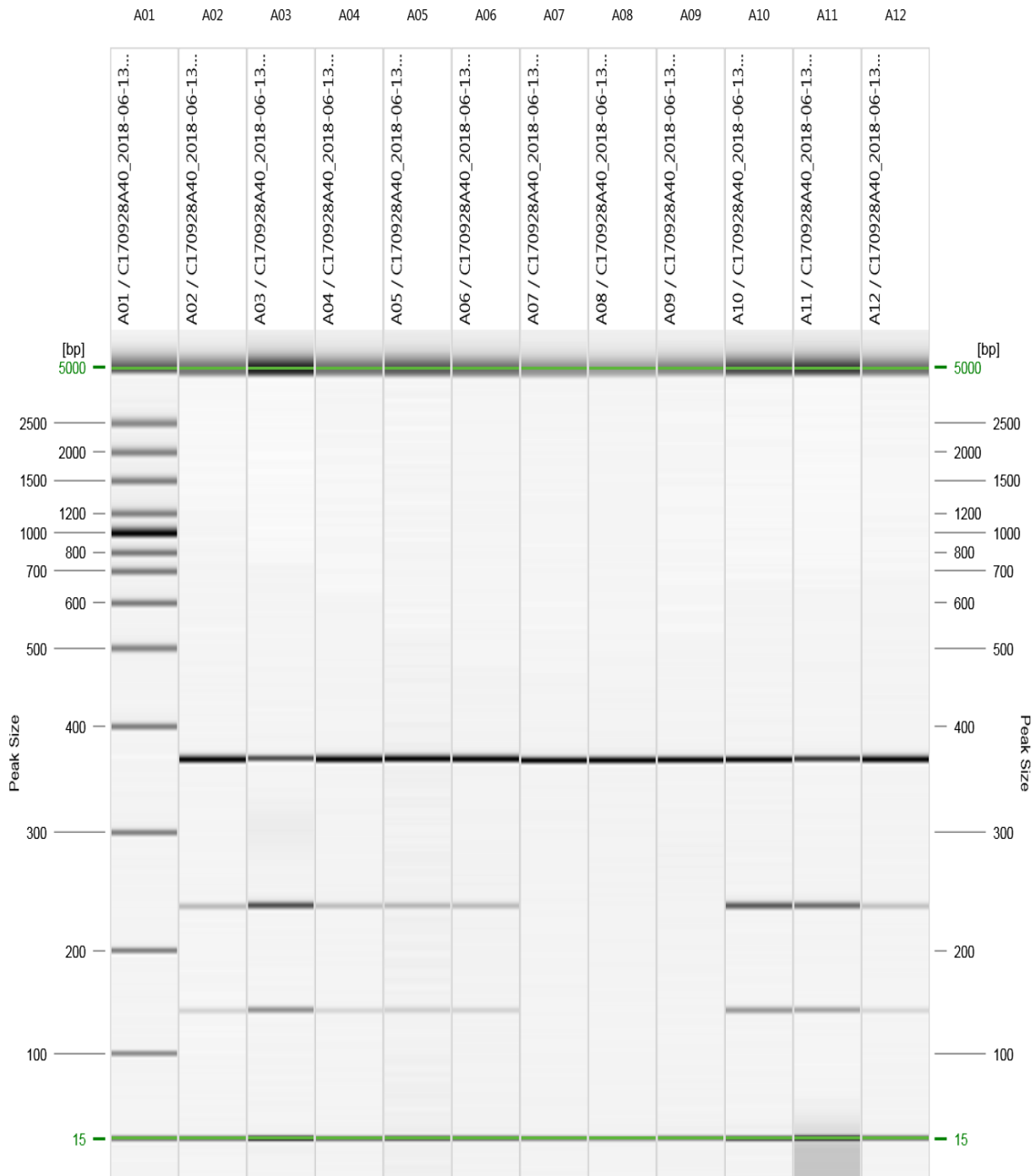


Figure: 1

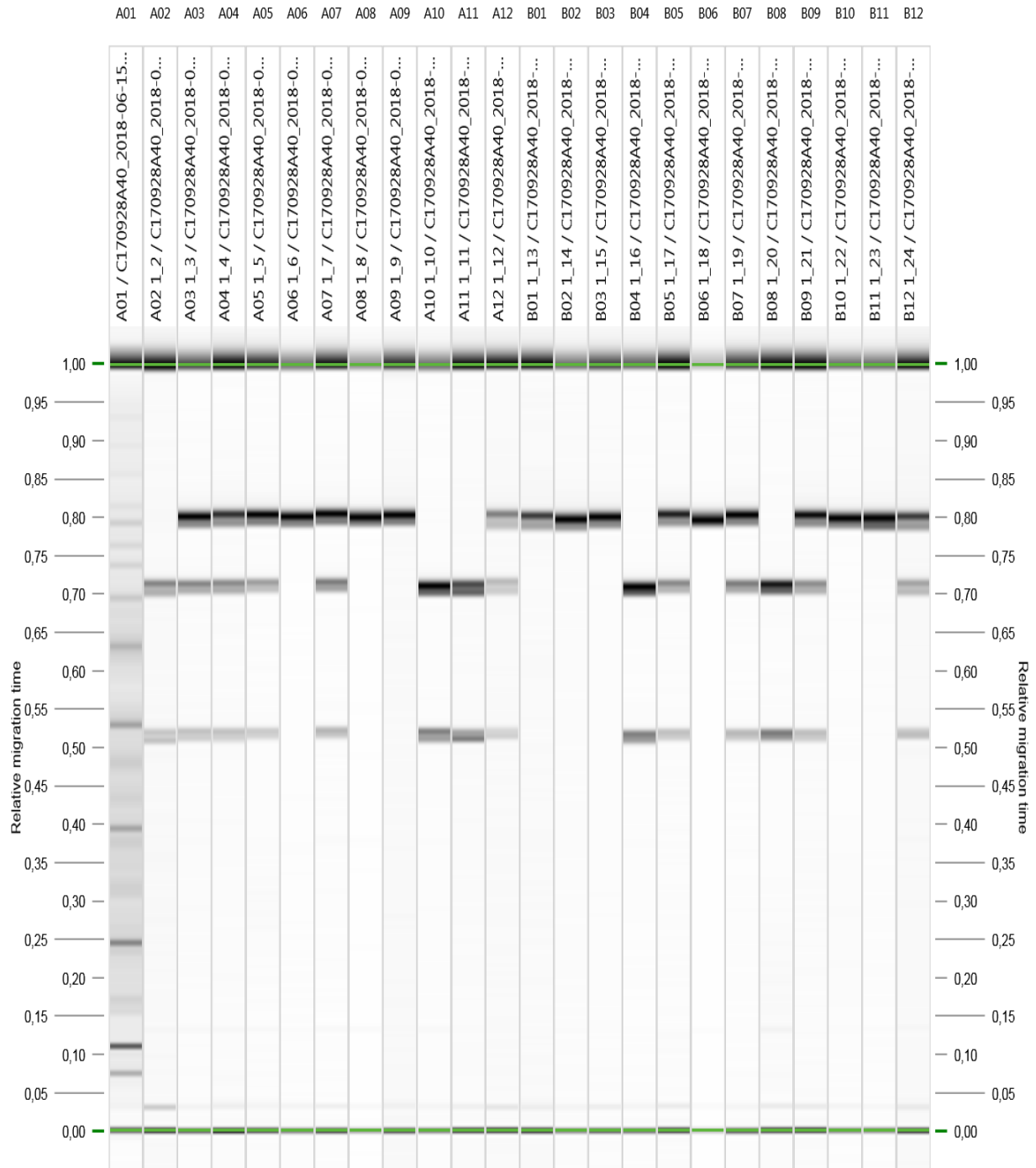


Figure: 1

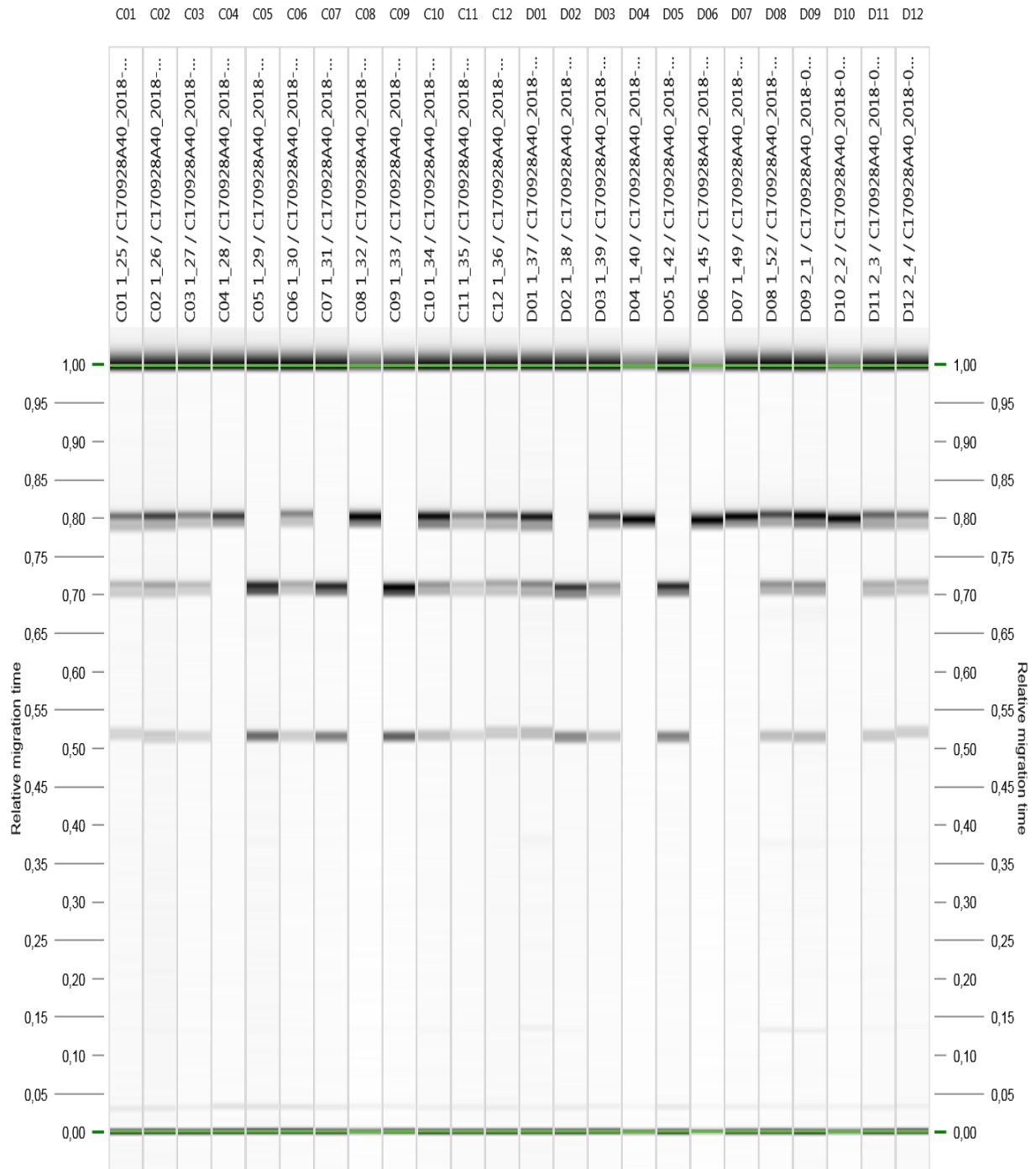


Figure: 2

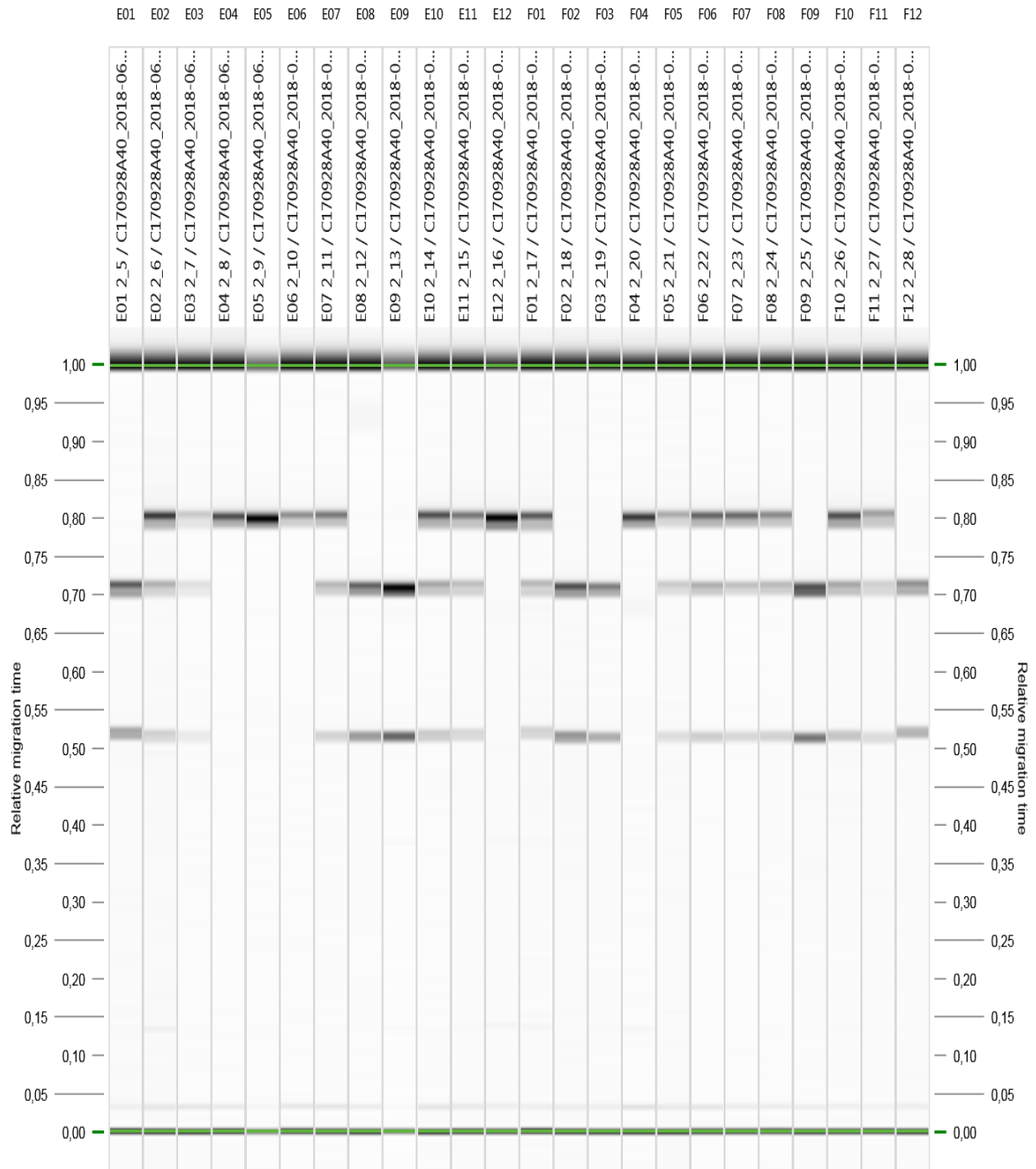


Figure: 3

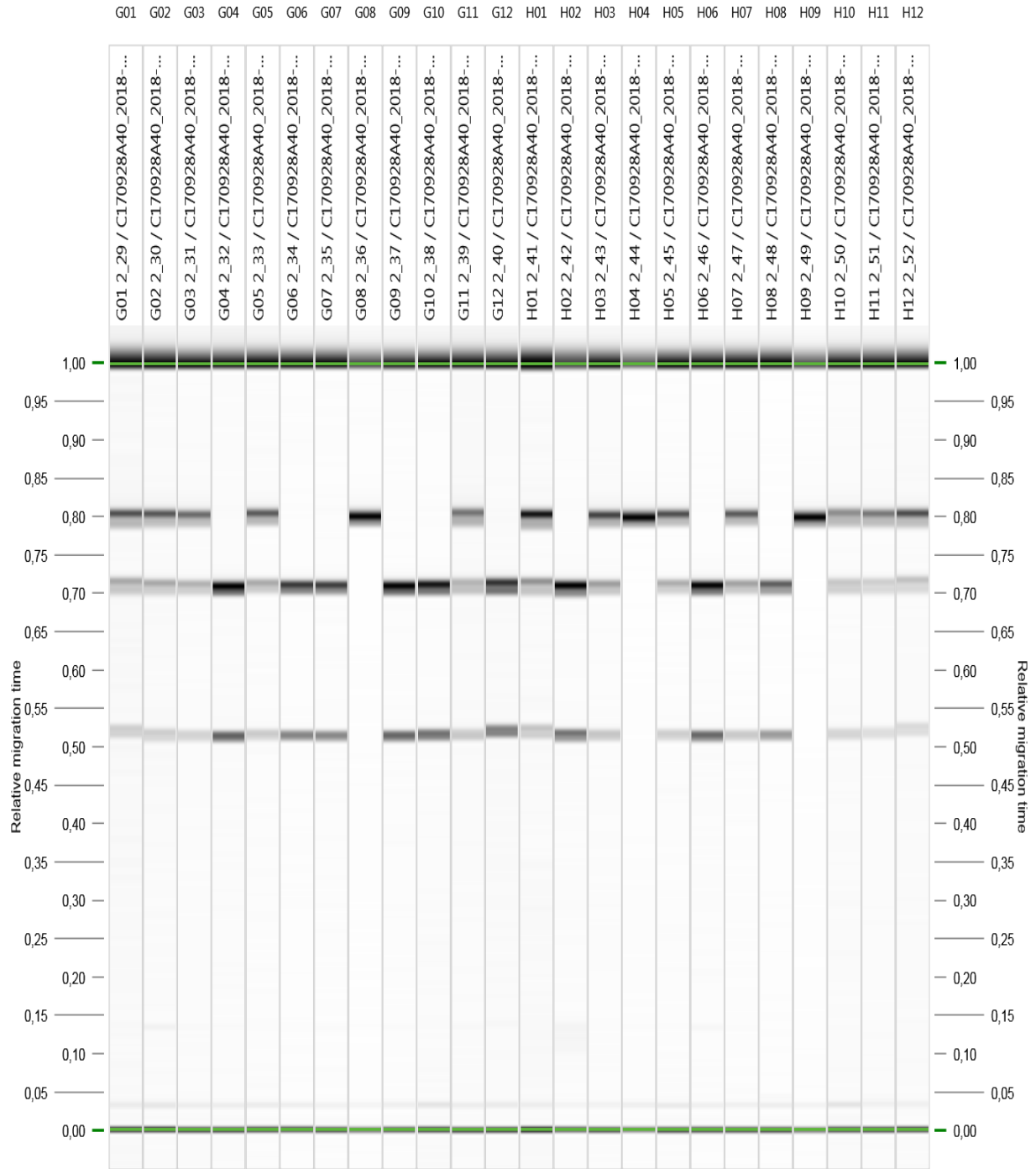


Figure: 4

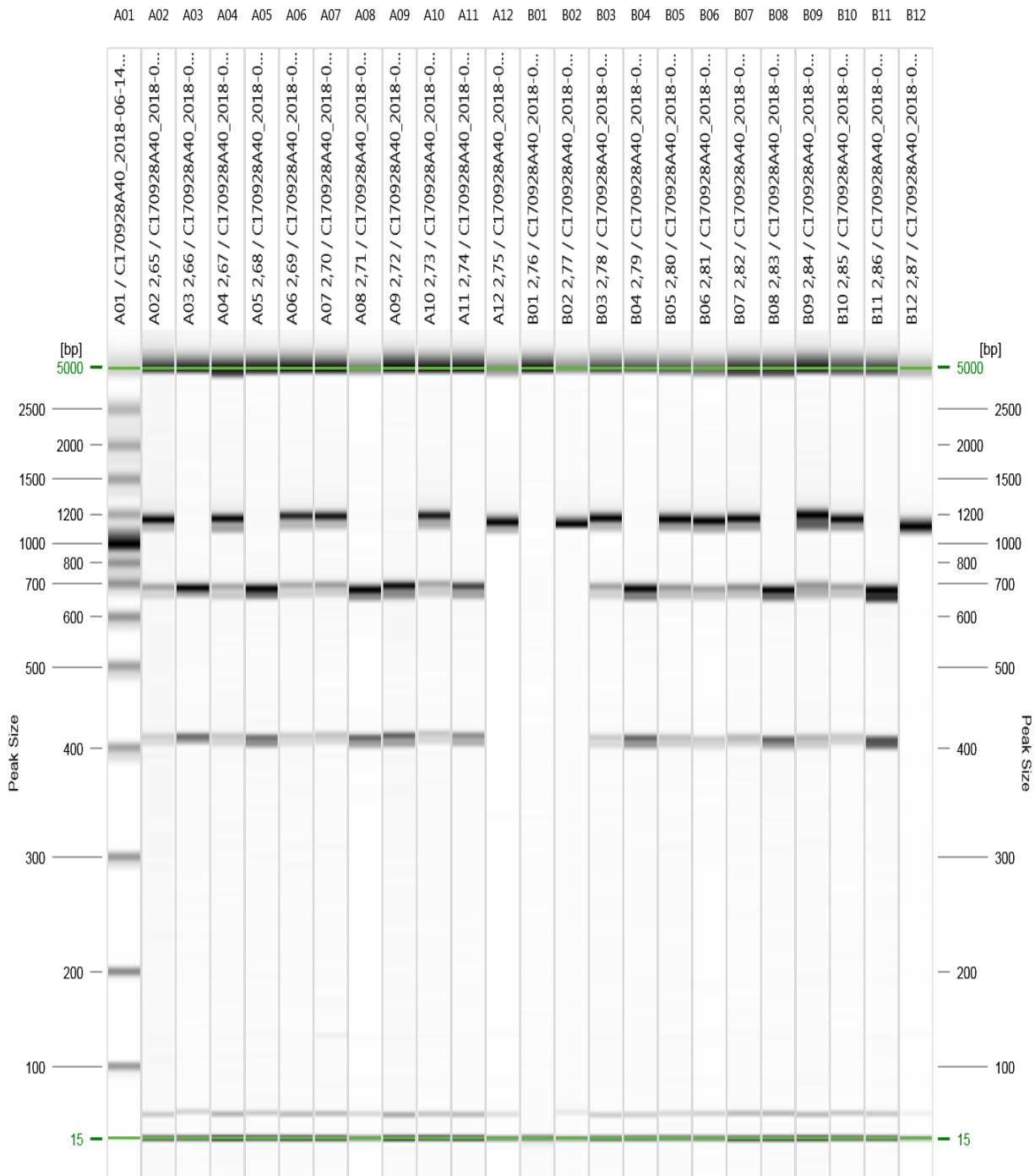


Figure: 1

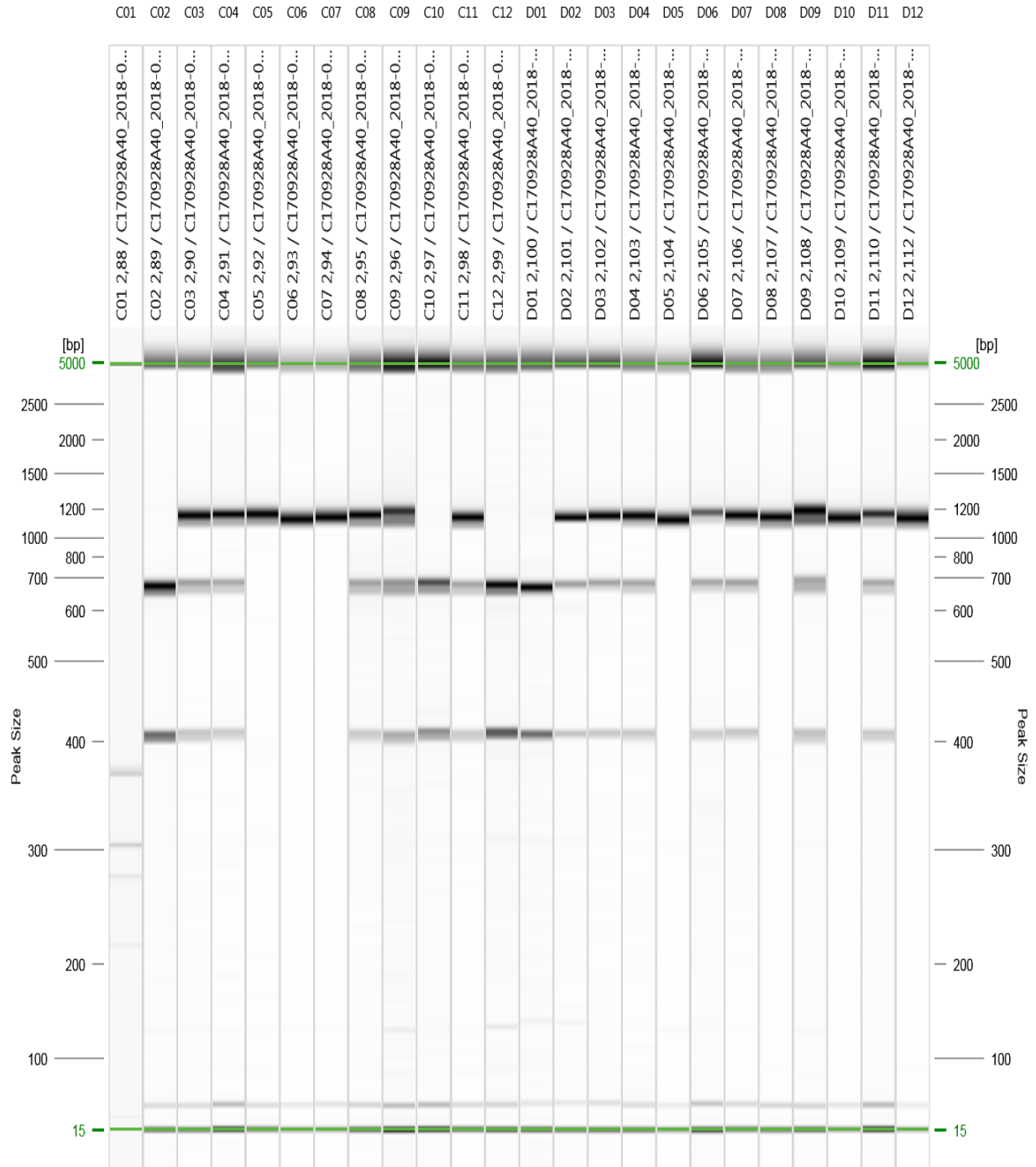


Figure: 2

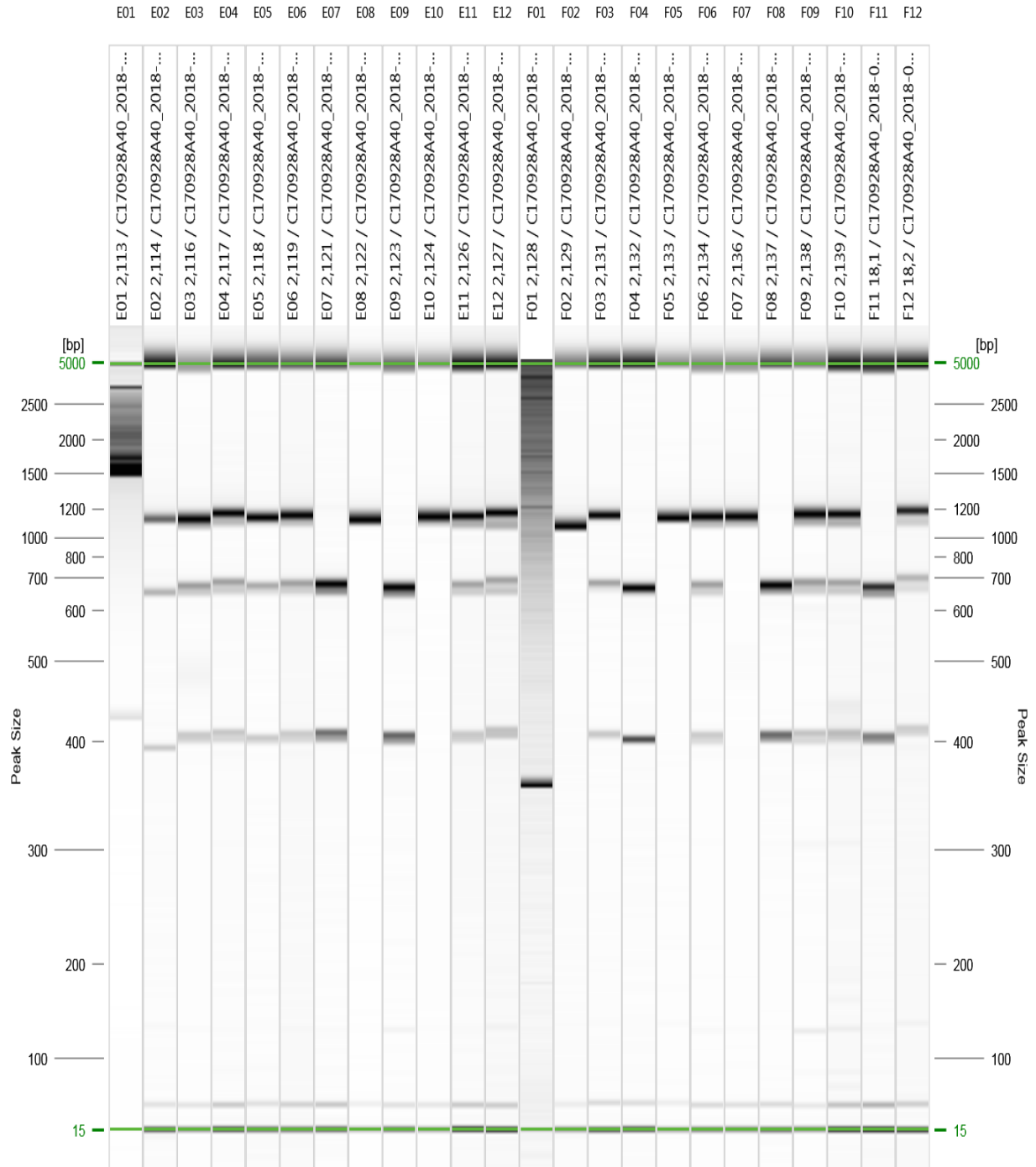


Figure: 3

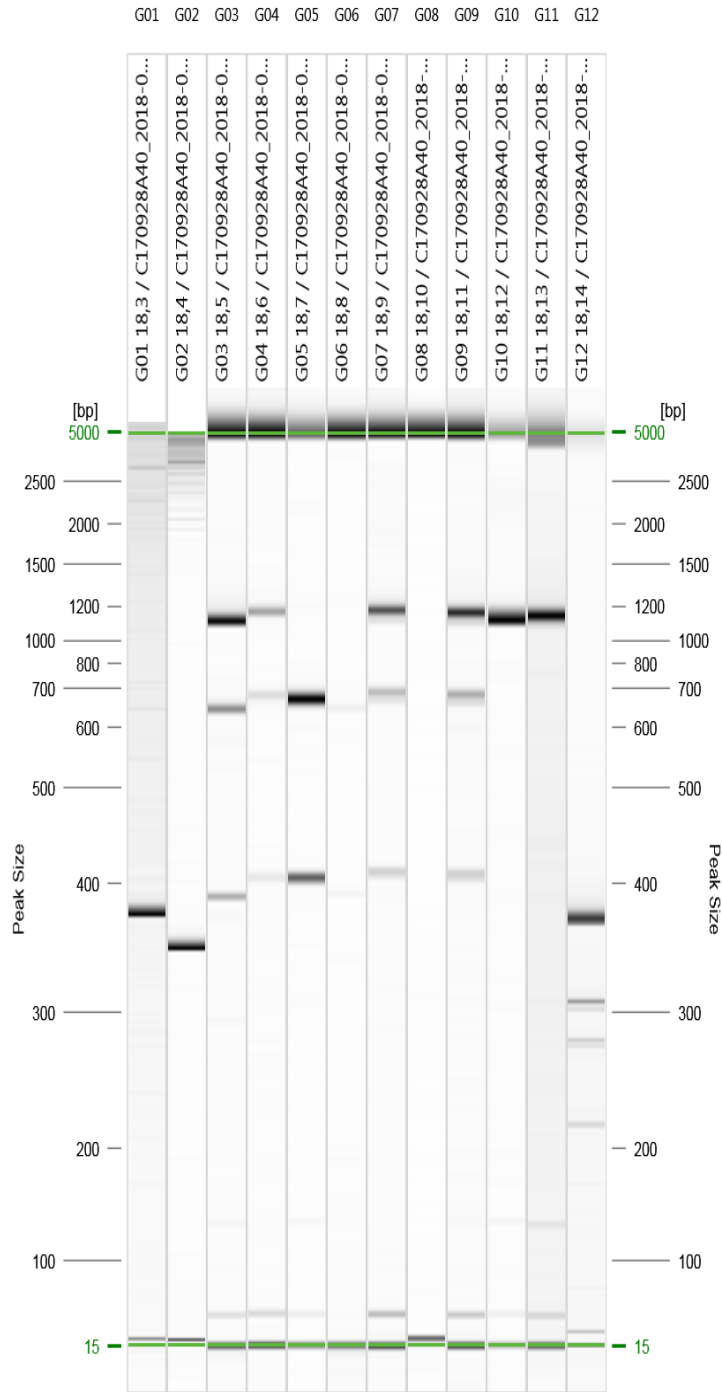


Figure: 4

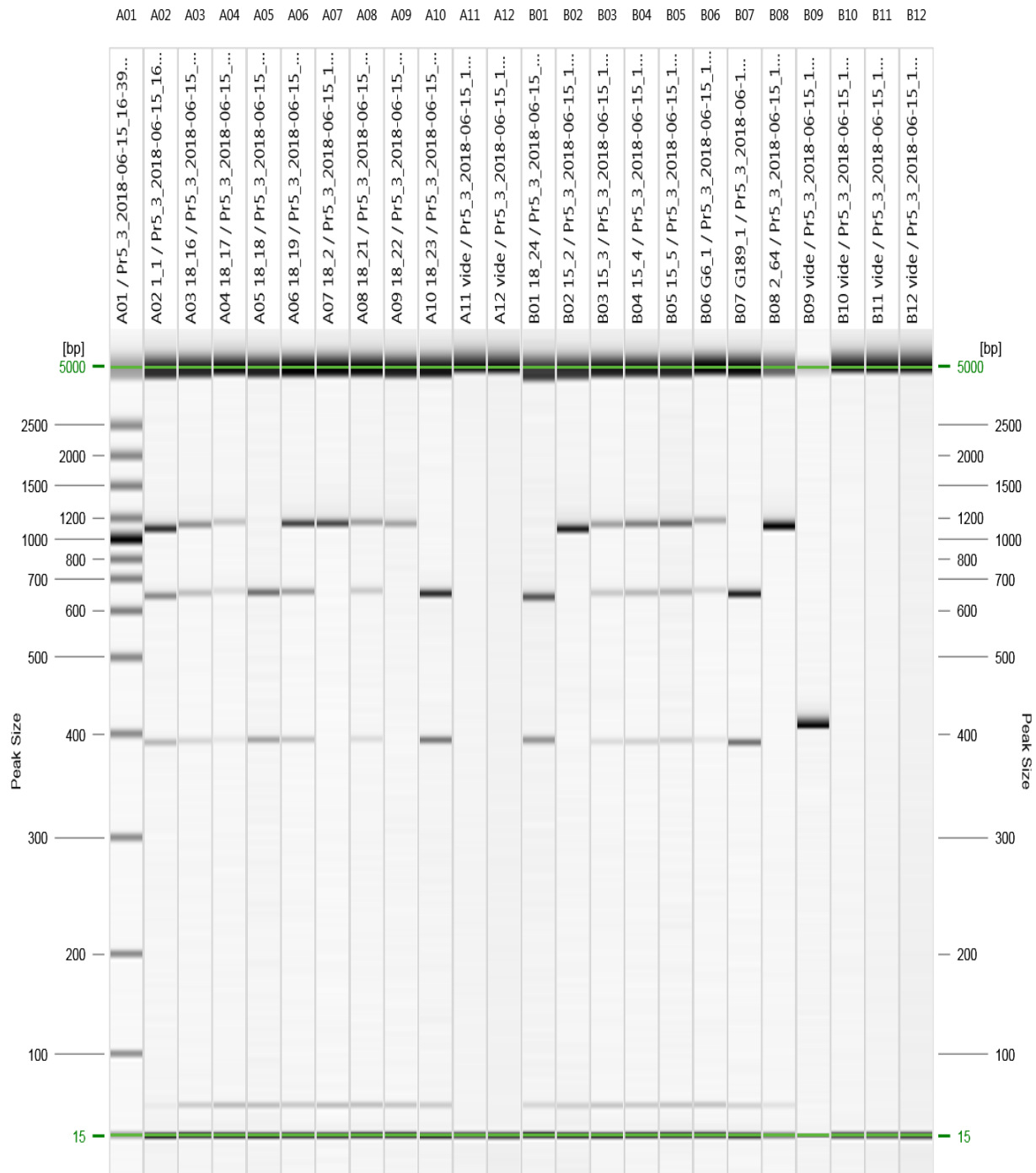


Figure: 1

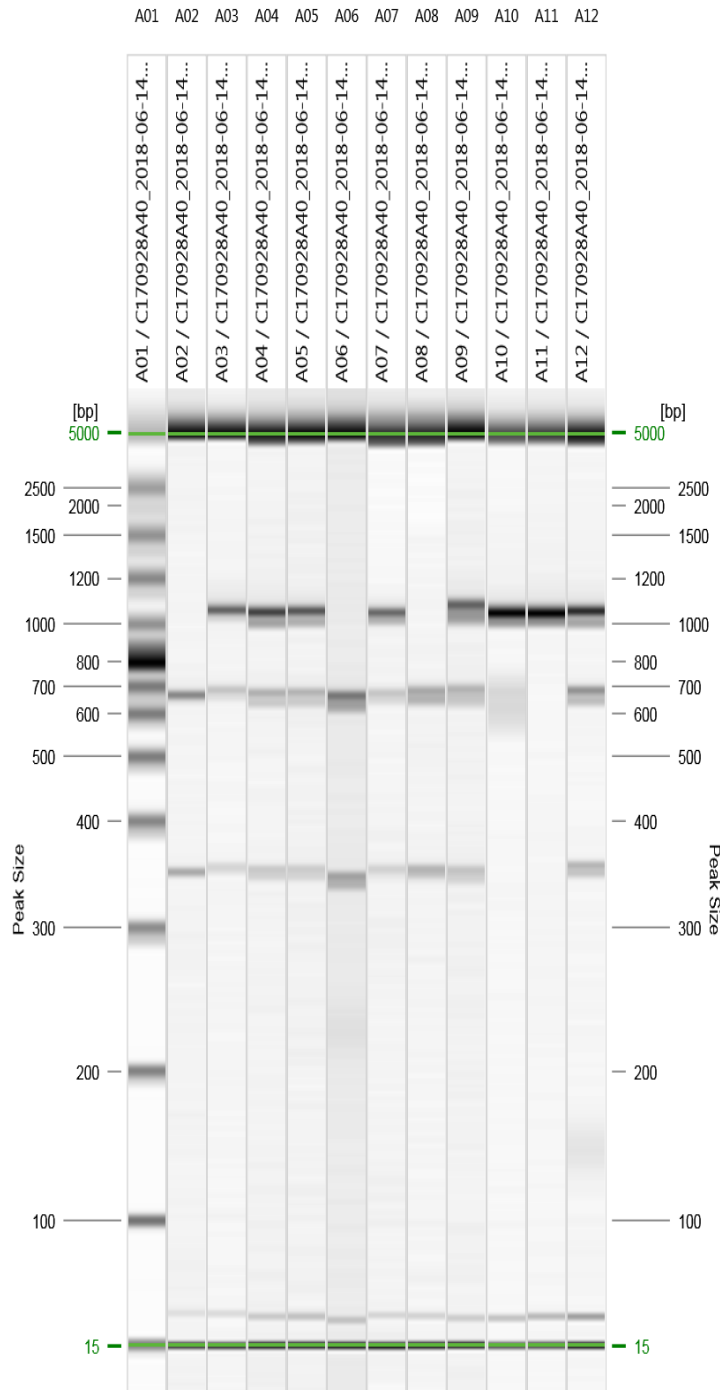


Figure: 1

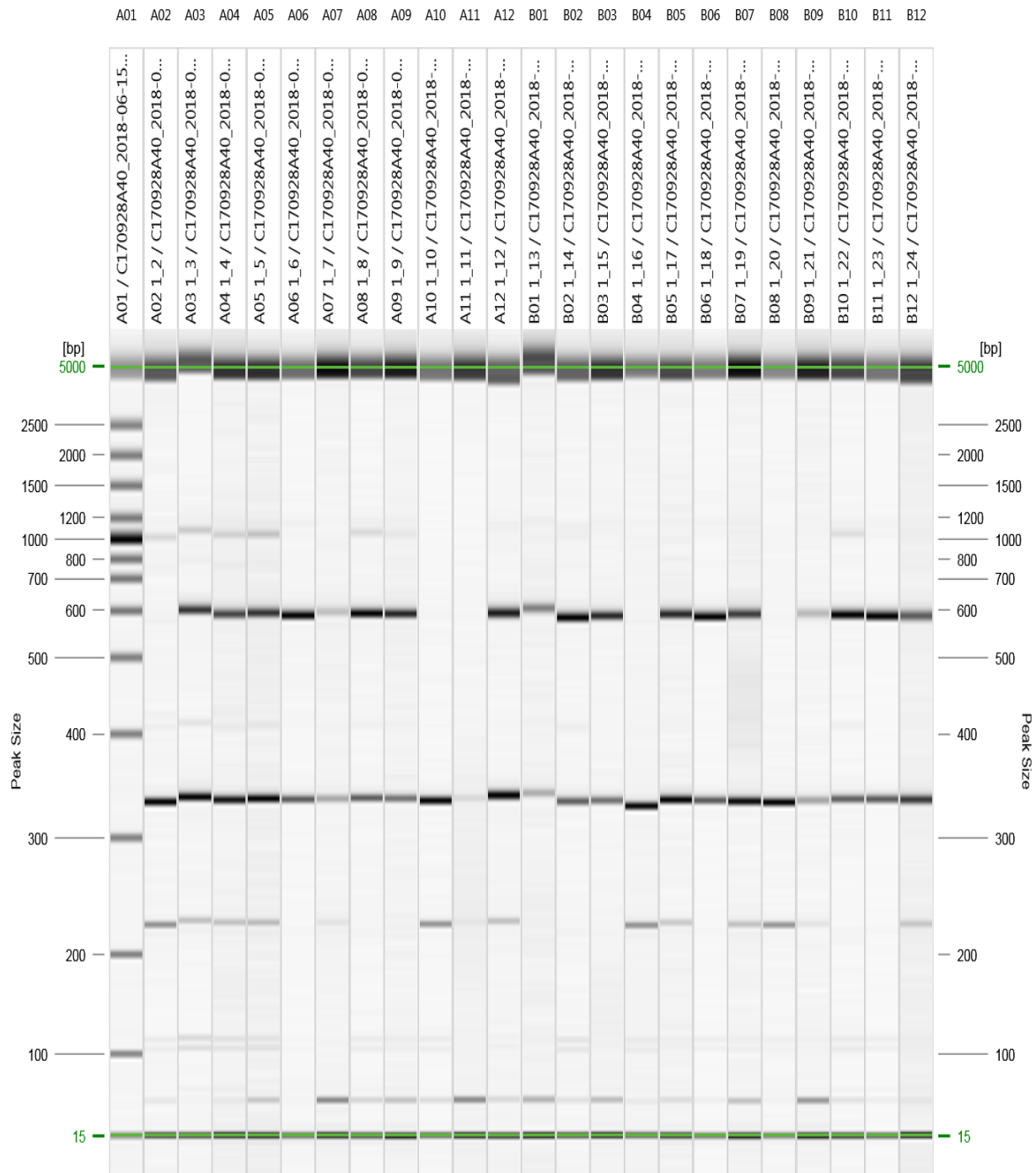


Figure: 1

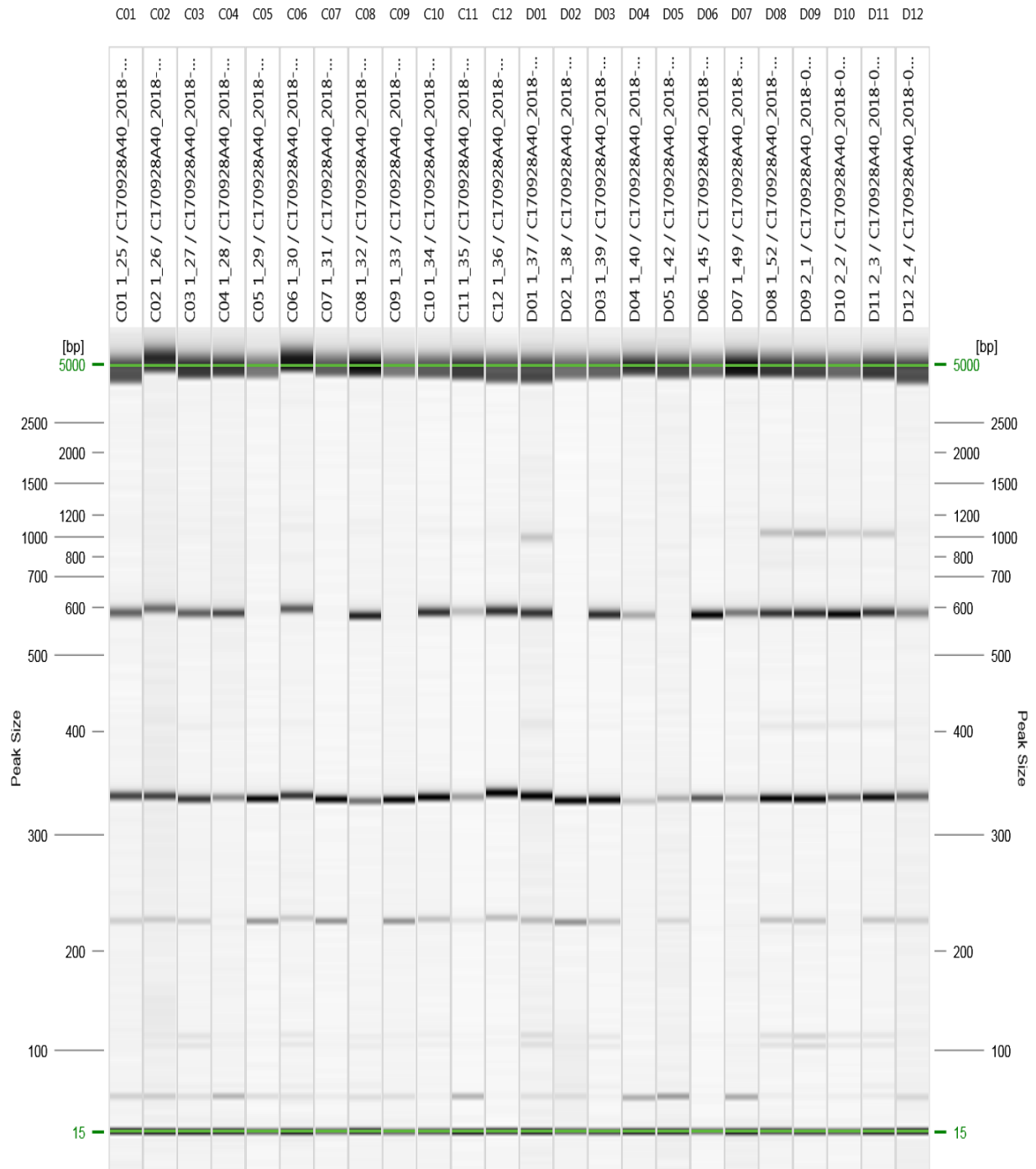


Figure: 2

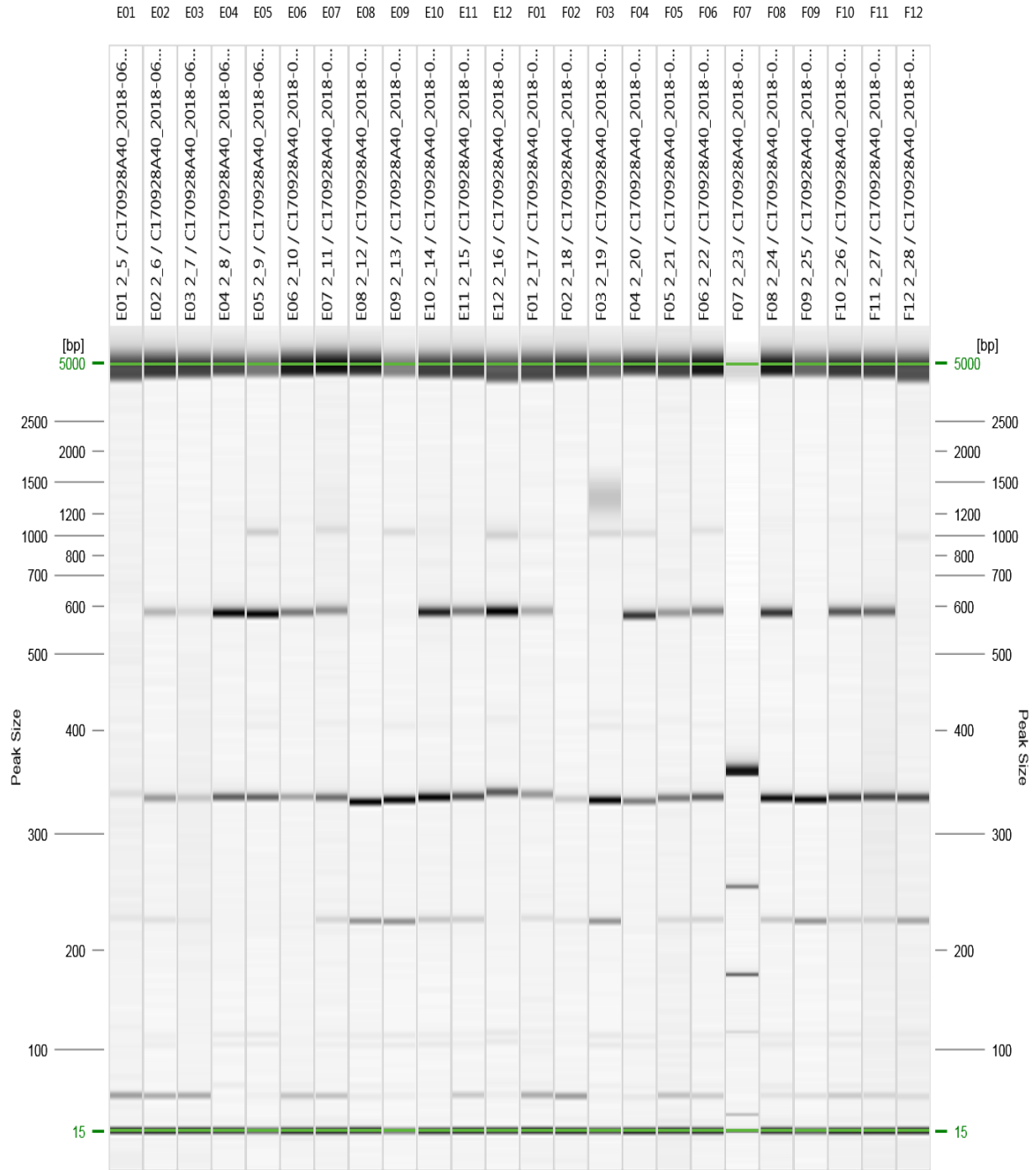


Figure: 3

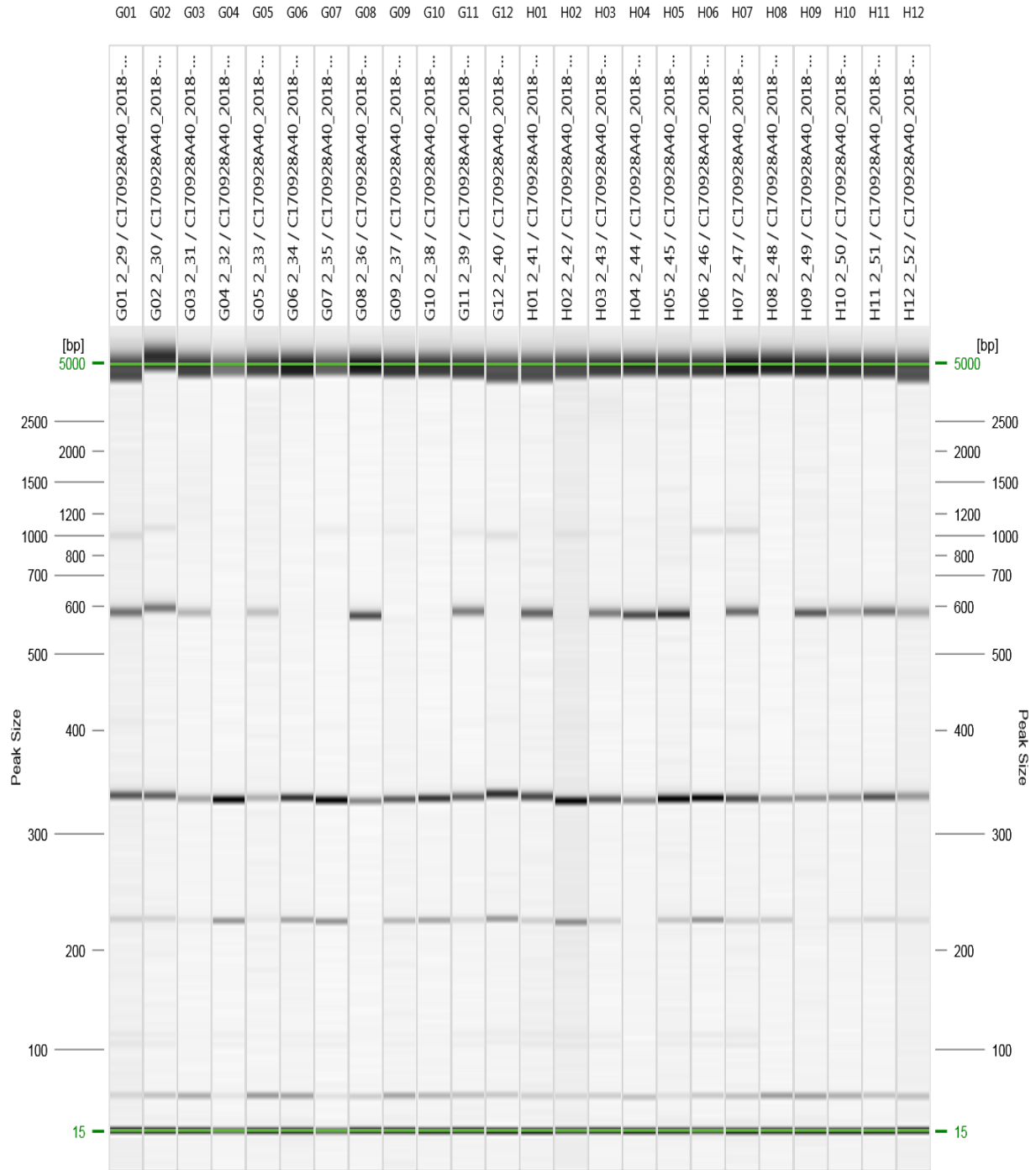


Figure: 4

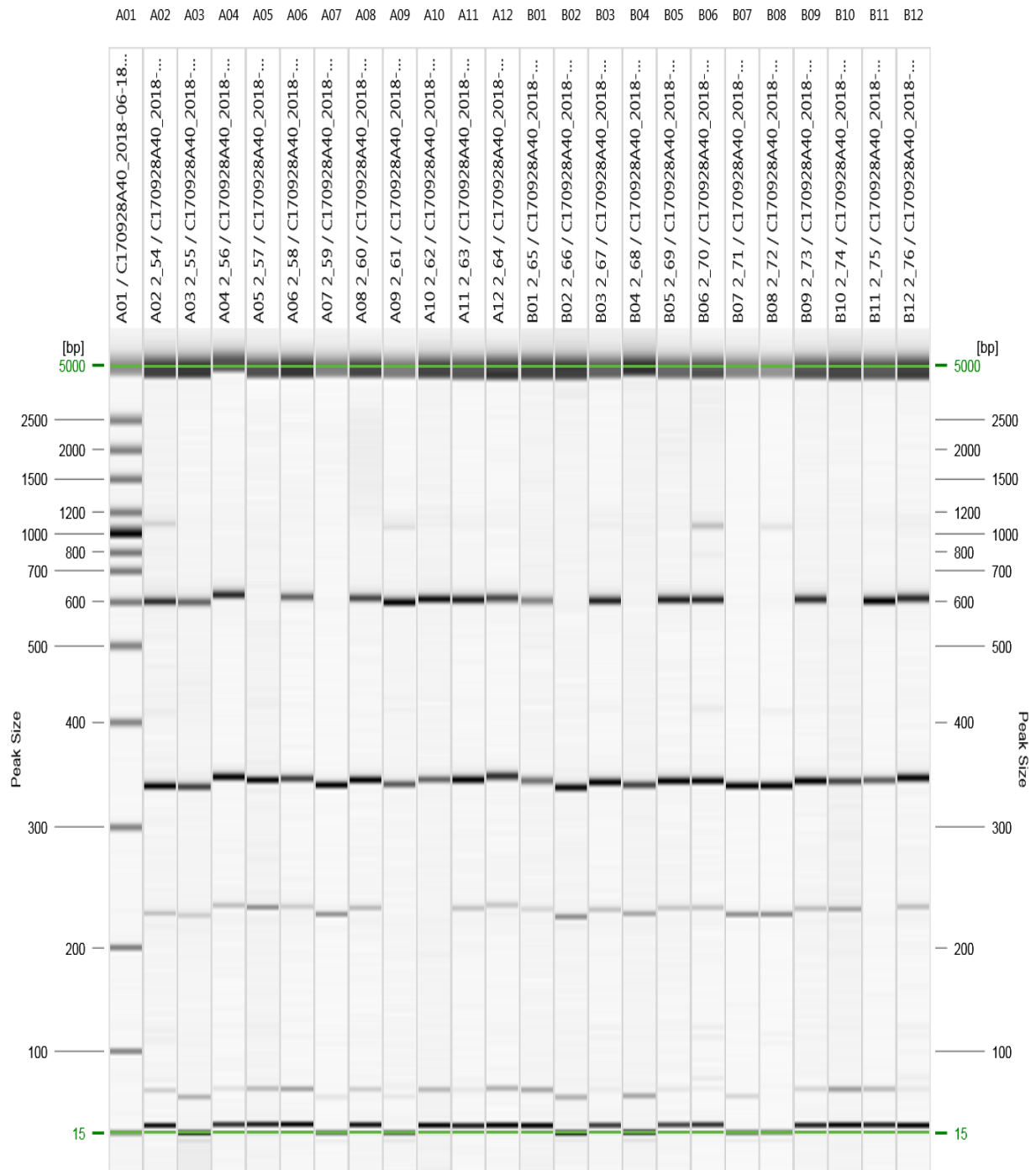


Figure: 1

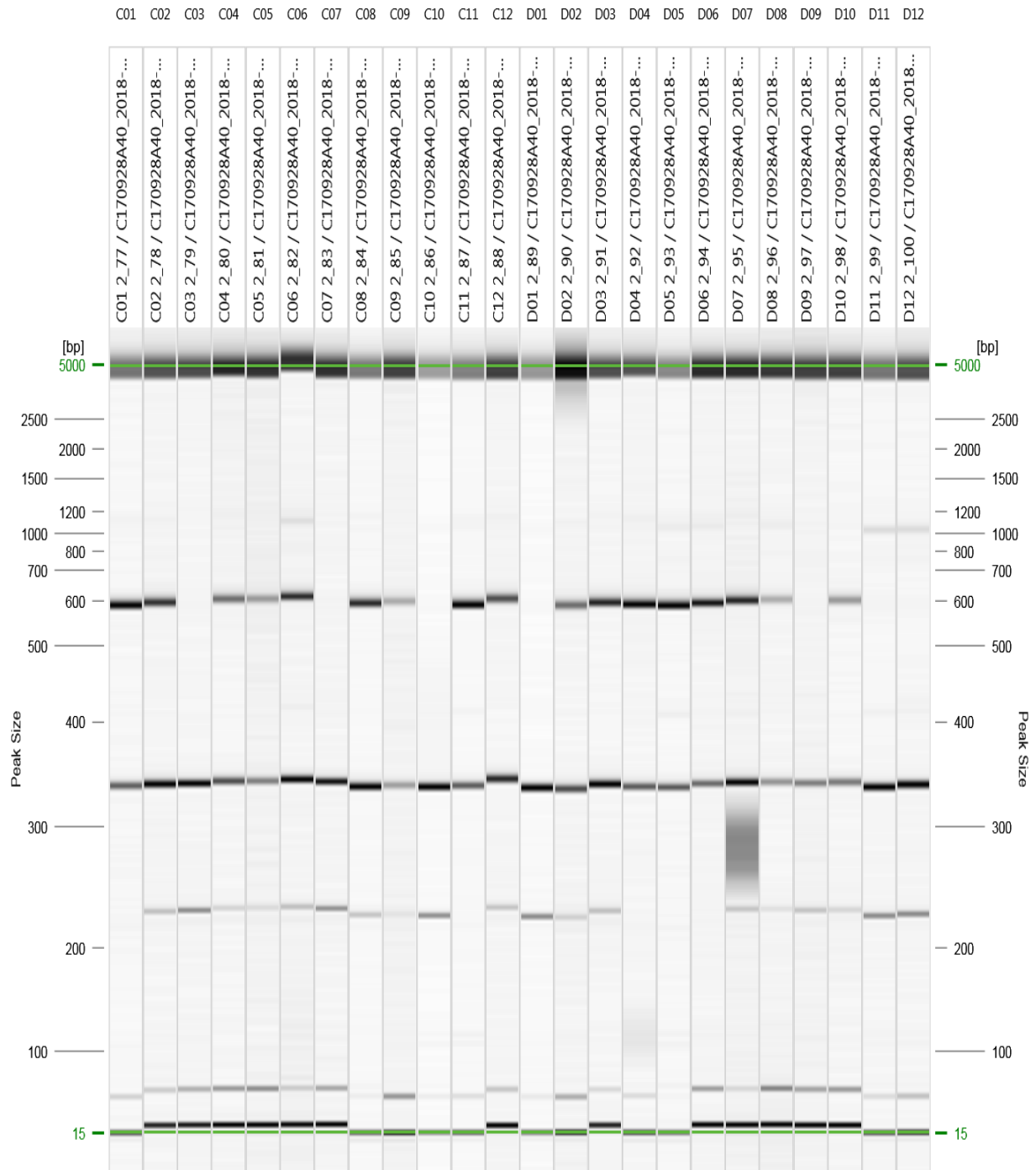


Figure: 2

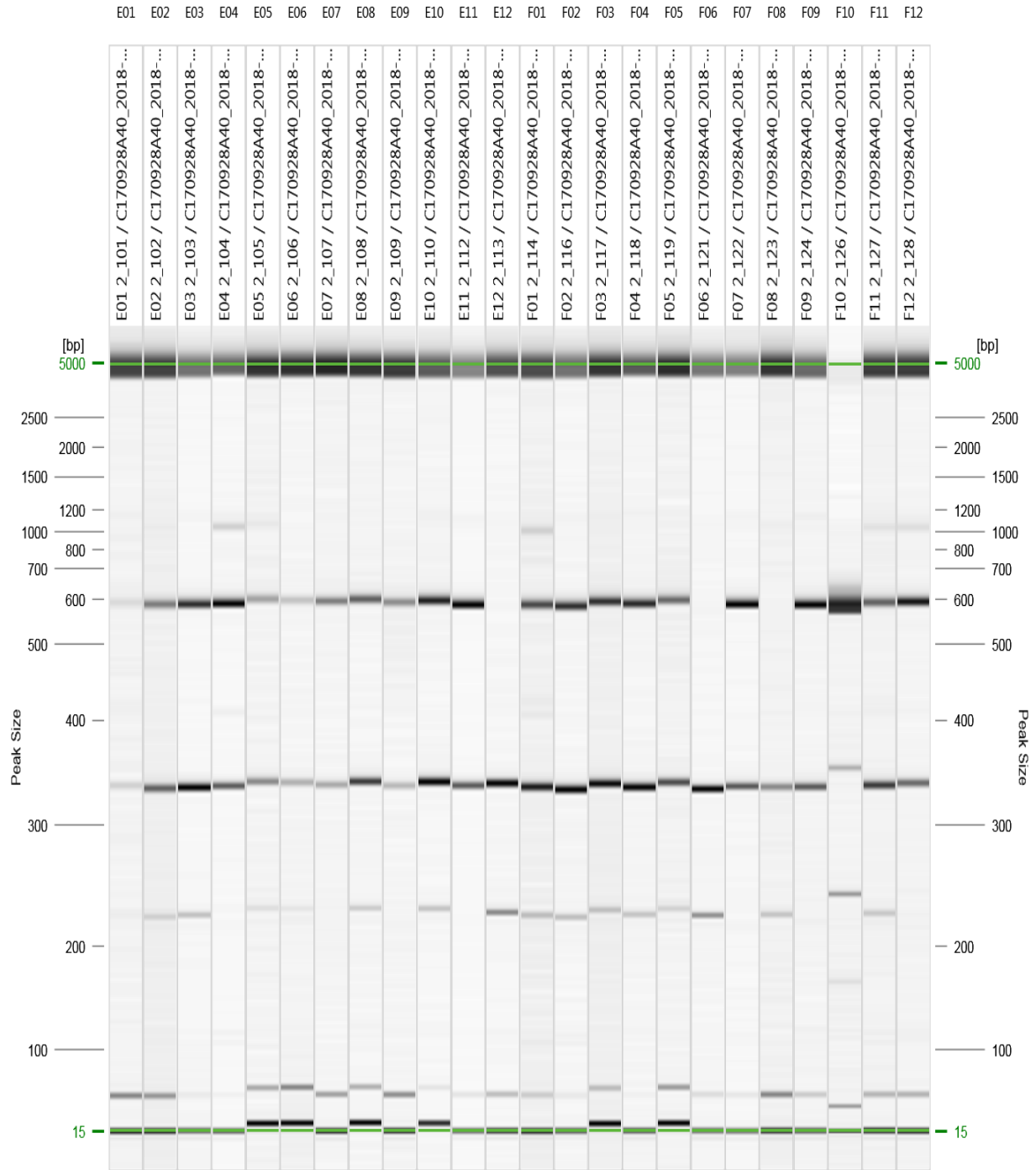


Figure: 3

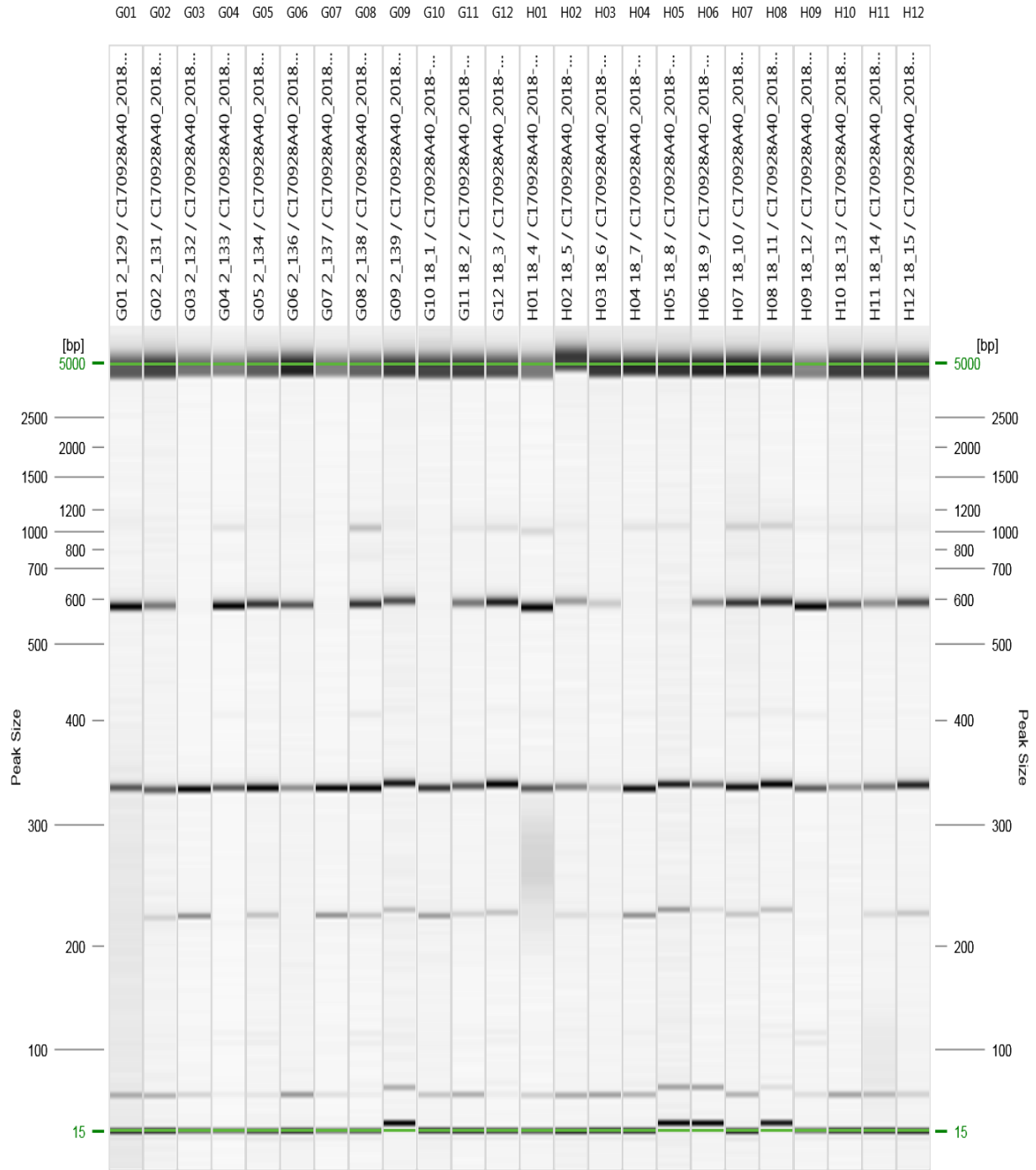


Figure: 4

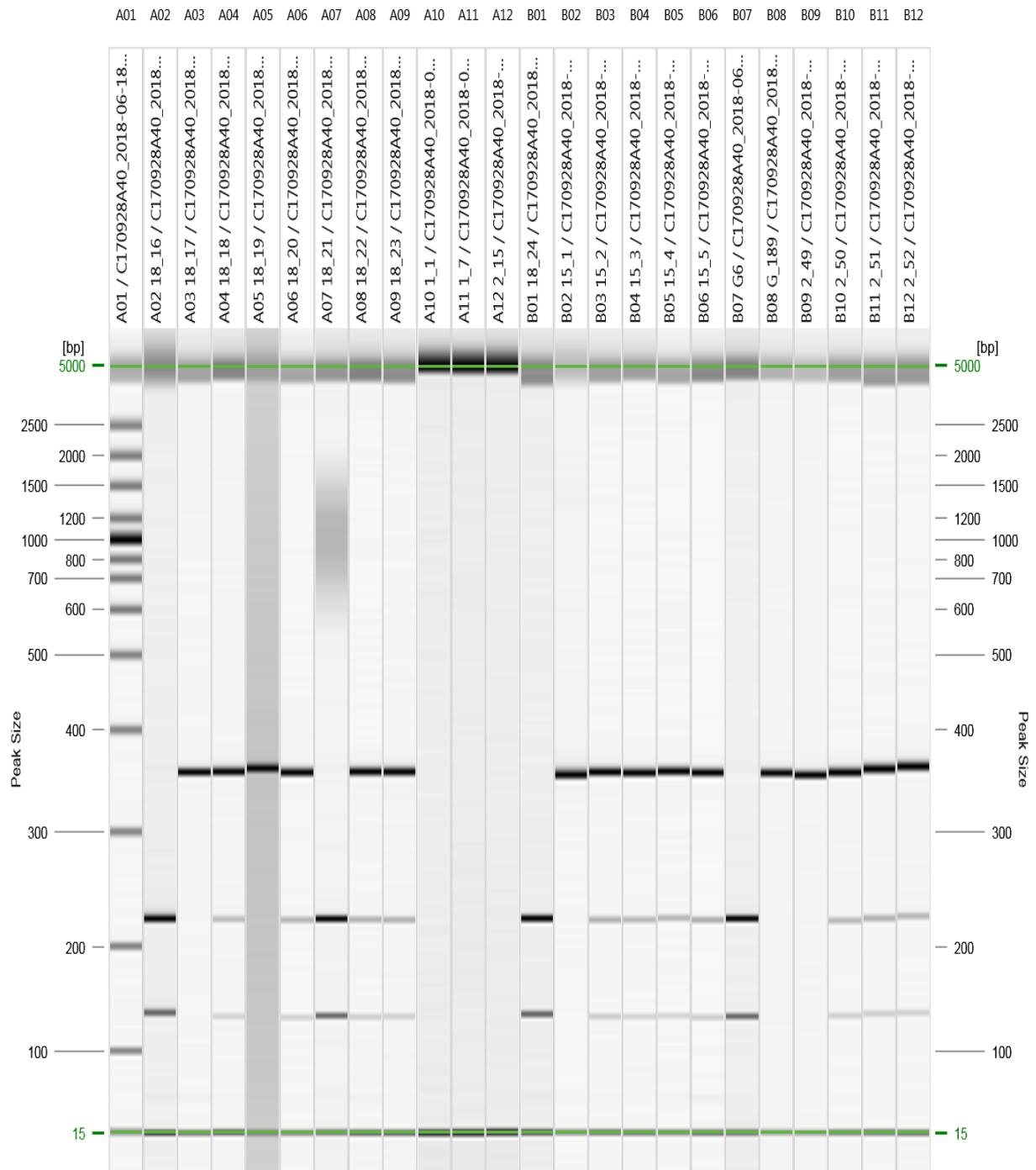


Figure: 1

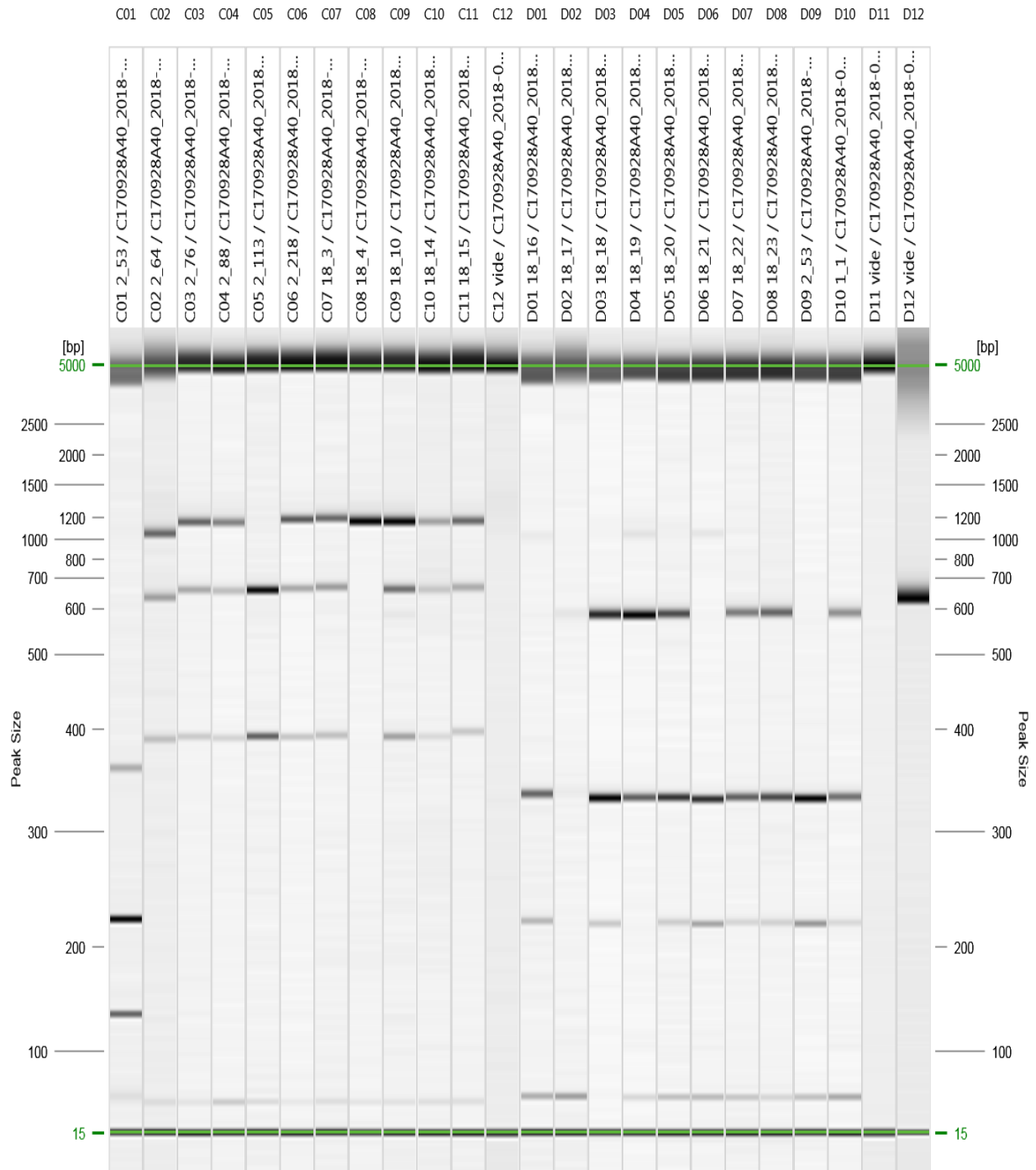


Figure: 2

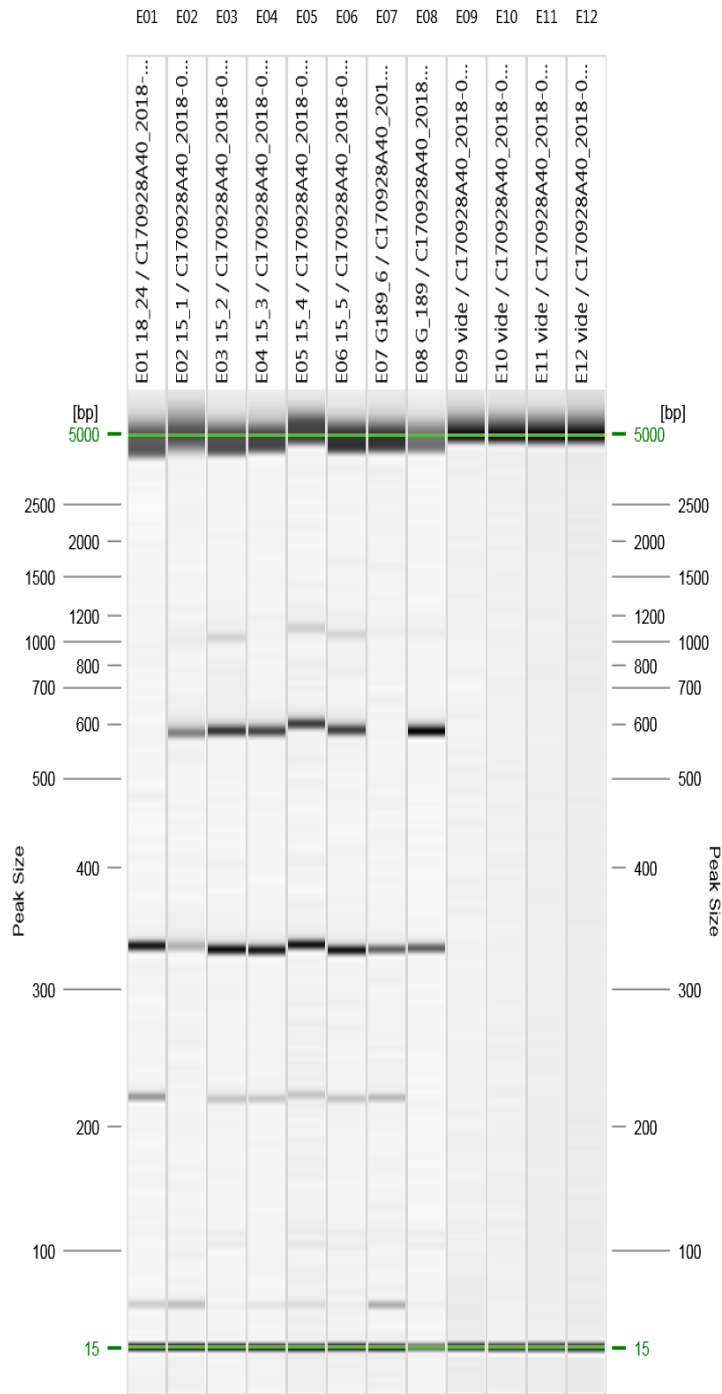


Figure: 3

**PHỤ LỤC 4. SƠ ĐỒ BỐ TRÍ THÍ NGHIỆM ĐỒNG RUỘNG
VỤ XUÂN NĂM 2019, TẠI XÃ ĐÔNG LA, HOÀI ĐỨC, HÀ NỘI**

B1		B2		B3		B4		B5		
0,5	1	0,5	1	0,5	1	0,5	1	0,5	1	0,5
	G6_29		G6_25		G189_24		G6_32		G189_27	
	G6_20		G189_43		G189_46		G6_34		G6_45	
	G6_44		G6_15		G189_19		G189_26		G189_9	
	G189_13		G189_47		G189_11		G6_17		G189_4	
	G6_22		G6_21		G6_4		G6_6		G6_48	
	G6_10		G6_39		G189_18		G189_49		G189_36	
	G189_34		G6_35		G6_38		G189_8		G6_42	
	G189_16		G6_19		G6_43		G6_11		G189_20	
	G6_31		G6_26		G189_10		G189_37		G189_38	
	G6_13		G189_6		G6_36		G6_41		G189_25	
	G6_1		G189_33		G189_39		G6_33		G6_3	
	G189_15		G189_5		G189_1		G6_23		G6_49	
	G189_14		G6_27		G189_42		G189_31		G189_2	
	G189_32		G6_30		G6_37		G189_40		G6_7	
	G189_29		G6_28		G6_5		G6_0		G189_23	
	G189_3		G6_47		G6_46		G6_16		G189_44	
	G189_21		G189_28		G6_18		G6_40		G189_35	
	G189_48		G189_12		G6_14		G189_7		G189_0	
	G189_30		G6_12		G6_2		G6_24		G6_9	
	G189_22		G189_41		G189_17		G6_8		G189_45	