

NGHIÊN CỨU, KHAI THÁC BIẾN DỊ SNP VÀ INDEL GENE PHỤC VỤ PHÁT TRIỂN CHỈ THỊ PHÂN TỬ TRONG CHỌN TẠO GIỐNG LÚA LIÊN QUAN ĐẾN NĂNG SUẤT VÀ CHẤT LƯỢNG

Khổng Ngân Giang¹, Võ Thị Minh Tuyền^{1,*}, Nguyễn Thị Hảo¹, Đoàn Văn Sơn¹,
Trần Thị Huệ Hương², Nguyễn Trường Giang³, Trần Mạnh Cường³, Lê Đức Thảo¹

¹ Viện Di truyền Nông nghiệp

² Viện Khoa học Nông nghiệp Việt Nam

³ Bộ Nông nghiệp và Môi trường

*Email: minhhtuyenagi@gmail.com

TÓM TẮT

Nghiên cứu được thực hiện nhằm khai thác các biến dị SNP và INDEL phục vụ phát triển chỉ thị phân tử ở các dòng lúa đột biến triển vọng. Dữ liệu giải trình tự hệ gene của ba dòng lúa đột biến A64, A68, A83 và giống đối chứng D3 được sử dụng để phân tích biến dị trên 186 gene liên quan đến sinh trưởng, năng suất và chất lượng lúa. Các biến dị SNP và INDEL nằm trong vùng trình tự gene được trích xuất, sàng lọc và so sánh giữa các dòng đột biến với giống đối chứng nhằm xác định các locus đa hình tiềm năng phục vụ phát triển chỉ thị phân tử. Kết quả phân tích xác định được tổng cộng 10.614 biến thể SNP và INDEL trong tập hợp các gene khảo sát, bao gồm: 3.479 biến thể ở dòng A64; 3.877 biến thể ở dòng A68 và 3.258 biến thể ở dòng A83. Nhiều gene liên quan đến cấu trúc bông, kiến trúc cây và kích thước hạt ghi nhận mức đa hình cao, tiêu biểu như: *ANK1*, *ANK2*, *GN2*, *APO1/SCM2*, *DEP1*, *IPA1/OsSPL14*, *GW2* và *GS5*. Từ các biến dị đã xác định, nghiên cứu đã lựa chọn và thiết kế 42 chỉ thị INDEL ứng viên phục vụ nhận diện kiểu gene và phát triển chỉ thị phân tử. Kết quả nghiên cứu cung cấp nguồn dữ liệu biến dị di truyền có giá trị và các chỉ thị phân tử tiềm năng phục vụ đánh giá nguồn gene và hỗ trợ chọn giống lúa bằng chỉ thị phân tử trong tương lai.

Từ khóa: Lúa đột biến, SNP, INDEL, chỉ thị phân tử, chọn giống lúa, năng suất lúa.

1. ĐẶT VẤN ĐỀ

Lúa (*Oryza sativa* L.) là cây lương thực quan trọng, cung cấp nguồn lương thực chính cho hơn một nửa dân số thế giới và giữ vai trò đặc biệt quan trọng đối với an ninh lương thực tại nhiều quốc gia châu Á [1]. Trong bối cảnh dân số gia tăng, diện tích đất nông nghiệp suy giảm và tác động ngày càng rõ rệt của biến đổi khí hậu, việc phát triển các giống lúa có năng suất cao, chất lượng tốt và thích nghi với điều kiện bất lợi đang trở thành yêu cầu cấp thiết [2]. Những năm gần đây, nhiều gene liên quan đến các tính trạng nông sinh học quan trọng như: Cấu trúc bông, số hạt trên bông, kích thước hạt, thời gian sinh trưởng và chất lượng gạo đã được xác định và ứng dụng trong nghiên cứu chọn giống lúa [3, 4]. Việc khai thác hiệu quả các biến dị di truyền trên các gene

này có ý nghĩa quan trọng trong phát hiện các allele có lợi và phát triển các công cụ hỗ trợ chọn giống hiện đại.

Sự phát triển nhanh của công nghệ giải trình tự thế hệ mới (Next Generation Sequencing - NGS) đã tạo điều kiện thuận lợi cho việc phân tích hệ gene và phát hiện biến dị di truyền ở quy mô lớn [5]. Trong đó, SNP (Single Nucleotide Polymorphism) và INDEL (Insertion/Deletion) là hai dạng biến dị phổ biến được sử dụng rộng rãi trong nghiên cứu đa dạng di truyền, lập bản đồ gene, phân tích chức năng gene và chọn giống bằng chỉ thị phân tử [6]. Các biến dị này có ưu điểm phân bố rộng khắp hệ gene, độ ổn định cao và có khả năng liên kết với nhiều tính trạng nông sinh học quan trọng [7]. Đặc biệt, các chỉ thị INDEL có thể được phát hiện bằng phản ứng PCR thông thường kết hợp điện di agarose, giúp

giảm chi phí phân tích và phù hợp với điều kiện của nhiều phòng thí nghiệm chọn giống [8].

Trong chọn giống lúa, việc khai thác dữ liệu SNP và INDEL từ giải trình tự hệ gene đang được ứng dụng ngày càng rộng rãi nhằm phát triển các chỉ thị phân tử phục vụ đánh giá nguồn gene và hỗ trợ chọn giống [9]. Nhiều nghiên cứu đã sử dụng dữ liệu hệ gene để phát triển các chỉ thị phân tử liên kết với các tính trạng như năng suất, chống chịu và chất lượng gạo [10, 11]. Tuy nhiên, phần lớn các nghiên cứu tập trung vào khai thác biến dị ở các giống hoặc tập đoàn nguồn gene quy mô lớn, trong khi việc phân tích có định hướng các biến dị SNP và INDEL trên các gene liên quan trực tiếp đến năng suất và chất lượng ở các dòng lúa đột biến triển vọng vẫn còn hạn chế, đặc biệt trong điều kiện nghiên cứu tại Việt Nam. Điều này làm hạn chế việc khai thác hiệu quả nguồn biến dị phục vụ phát triển các chỉ thị phân tử ứng dụng trong chọn giống.

Các dòng lúa đột biến là nguồn vật liệu có giá trị trong nghiên cứu di truyền và chọn giống do có khả năng tích lũy nhiều biến dị liên quan đến các tính trạng nông sinh học quan trọng [12]. Việc phân tích dữ liệu biến dị trên các gene ứng viên ở các dòng lúa đột biến không chỉ góp phần làm rõ nguồn đa dạng di truyền mà còn tạo cơ sở cho việc phát triển các chỉ thị phân tử phục vụ chọn giống lúa bằng chỉ thị phân tử trong tương lai. Ba dòng lúa đột biến A64, A68 và A83 được lựa chọn làm vật liệu trong nghiên cứu này là các dòng triển vọng được chọn lọc từ quần thể đột biến của giống ST5, vẫn giữ được ưu điểm về chất lượng tương tự giống ST5 nhưng đã được cải tiến về năng suất và thời gian sinh trưởng so với giống gốc ban đầu, đồng thời đã được giải trình tự hệ gene, tạo điều kiện thuận lợi cho việc khai thác các biến dị di truyền phục vụ phát triển chỉ thị phân tử.

Trong nghiên cứu này, dữ liệu giải trình tự hệ gene của các dòng lúa đột biến A64, A68, A83 và giống đối chứng D3 được sử dụng để phân tích

các biến dị SNP và INDEL trên 186 gene liên quan đến sinh trưởng, năng suất và chất lượng lúa. Trên cơ sở đó, nghiên cứu tiến hành lựa chọn các biến dị tiềm năng và phát triển các chỉ thị INDEL ứng viên phục vụ nhận diện kiểu gene và hỗ trợ chọn giống lúa.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu nghiên cứu

Nghiên cứu sử dụng dữ liệu giải trình tự hệ gene của ba dòng lúa đột biến triển vọng A64, A68, A83 (các dòng lúa được chọn lọc trong quần thể đột biến giống Sóc Trăng 5 - ST5) và giống đối chứng D3 (giống ban đầu ST5). Các dòng vật liệu được lựa chọn trên cơ sở khả năng sinh trưởng và tiềm năng ứng dụng trong nghiên cứu chọn giống lúa.

Dữ liệu sử dụng trong nghiên cứu là bộ dữ liệu biến dị SNP và INDEL thu được từ kết quả giải trình tự hệ gene của các dòng nghiên cứu. Việc giải trình tự và phân tích dữ liệu sinh tin học được thực hiện theo phương pháp paired-end bởi Công ty TNHH Lobi Việt Nam (năm 2025), kết quả được bàn giao dưới dạng tệp Variant Call Format (VCF), chứa thông tin các biến dị SNP và INDEL của từng dòng vật liệu. Dữ liệu giải trình tự của các dòng A64, A68 và A83 đạt tỷ lệ reads đúng hàng với hệ gene tham chiếu trên 98%, với độ sâu giải trình tự trung bình lần lượt là 26,15×, 39,12× và 30,27×, đáp ứng yêu cầu cho phân tích biến dị SNP và INDEL. Bộ dữ liệu VCF được sử dụng làm nguồn dữ liệu đầu vào cho các bước trích xuất, sàng lọc và phân tích biến dị trong nghiên cứu này.

Danh sách 186 gene (Bảng 1) liên quan đến sinh trưởng, năng suất và chất lượng lúa được tổng hợp từ các công bố quốc tế và các cơ sở dữ liệu hệ gene lúa (MSU Rice Genome Annotation Project và RAP-DB). Các gene khảo sát bao gồm: Các nhóm gene liên quan đến cấu trúc bông, số hạt trên bông, chiều cao cây, thời gian sinh trưởng, kích thước hạt và chất lượng gạo.

Bảng 1. Danh sách 186 gene liên quan đến sinh trưởng, năng suất và chất lượng lúa được sử dụng trong nghiên cứu

TT	Locus_id MSU	Tên gene	TT	Locus_id MSU	Tên gene	TT	Locus_id MSU	Tên gene
1	LOC_Os01g03840	OsJAG	63	LOC_Os03g51330	GNA	125	LOC_Os06g41850	TGW6
2	LOC_Os01g08220	D18	64	LOC_Os03g51660	OsEPFL3	126	LOC_Os06g44100	OsgIHAT1
3	LOC_Os01g08700	OsGI	65	LOC_Os03g51690	OsH1 / Oskn1	127	LOC_Os06g44860	SPL10

KHOA HỌC & CÔNG NGHỆ

TT	Locus_id MSU	Tên gene	TT	Locus_id MSU	Tên gene	TT	Locus_id MSU	Tên gene
4	LOC_Os01g10110	Gn1a/OsCKX2	66	LOC_Os03g52239	OsBLH4	128	LOC_Os06g44970	OsPIN2 / LRA1
5	LOC_Os01g11040	TUTOU1	67	LOC_Os03g54084	PHYC	129	LOC_Os06g45460	AP01/SCM2
6	LOC_Os01g11940	OsFT-L1	68	LOC_Os03g54160	MADS14	130	LOC_Os06g46030	OsGIL2
7	LOC_Os01g13540	OsNPL3	69	LOC_Os03g54160	MADS14	131	LOC_Os06g46350	DEP3
8	LOC_Os01g33040	dbs1	70	LOC_Os03g54170	PAP2/MADS34	132	LOC_Os06g50340	FON1
9	LOC_Os01g40630	LOG1	71	LOC_Os03g55389	HD6/CK2	133	LOC_Os07g01820	OsMADS15 / DEP
10	LOC_Os01g47710	OsWOX7	72	LOC_Os03g57240	DST	134	LOC_Os07g03250	PLT8 / crl5
11	LOC_Os01g49160	RGN1	73	LOC_Os03g60430	IDS1	135	LOC_Os07g03770	D6 OSH15
12	LOC_Os01g51610	OsLFL1	74	LOC_Os03g61760	SPL6	136	LOC_Os07g04020	OsEPFL5
13	LOC_Os01g54270	D10	75	LOC_Os03g63970	GNP1/SDSFL1	137	LOC_Os07g04670	G1/ELE
14	LOC_Os01g54860	NOG1	76	LOC_Os03g64070	LOGL5	138	LOC_Os07g05900	PROG1
15	LOC_Os01g57580	OsHHLH069	77	LOC_Os04g01590	Ngf-1 (OsARG)	139	LOC_Os07g13170	SNB / SSH1
16	LOC_Os01g60900	OsEPFL9	78	LOC_Os04g28280	Awn-1	140	LOC_Os07g15770	Ghd7
17	LOC_Os01g61480	LAX1	79	LOC_Os04g33740	GIF1	141	LOC_Os07g41370	MADS18
18	LOC_Os01g66100	sd1	80	LOC_Os04g38720	OsNAC2/OsTIL1	142	LOC_Os07g42410	DEP2/EP2
19	LOC_Os01g67430	EG1/GY1	81	LOC_Os04g39430	GNS4/D11/CPB1	143	LOC_Os07g44210	LF2
20	LOC_Os01g69850	OsMADS51	82	LOC_Os04g41229	HBP1	144	LOC_Os07g47330	FZP / BFL1
21	LOC_Os01g71310	OsCKX4	83	LOC_Os04g45330	OsYABBY5	145	LOC_Os07g48660	OsFD7/OsZIP62
22	LOC_Os01g72090	SE13/OsHY2	84	LOC_Os04g46470	D17/OsCCD7	146	LOC_Os08g06100	COMT
23	LOC_Os01g72330	RR4	85	LOC_Os04g47890	MSF2/MOF1	147	LOC_Os08g06480	ASP1
24	LOC_Os02g02530	spp1	86	LOC_Os04g51000	AP02/RFL	148	LOC_Os08g07740	Ghd8/DTH8
25	LOC_Os02g05840	LC2/OsVIL3	87	LOC_Os04g52479	NAL1/qFLW4	149	LOC_Os08g12430	OsVIL1
26	LOC_Os02g05980	LRK1/qGY2-1	88	LOC_Os04g55230	FLO2	150	LOC_Os08g31470	PAY1
27	LOC_Os02g07030	OsGIL1	89	LOC_Os04g57720	RR6	151	LOC_Os08g37890	GAD1
28	LOC_Os02g14720	GW2	90	LOC_Os04g57720	RR6	152	LOC_Os08g39630	POH1
29	LOC_Os02g15950	LP/EP3	91	LOC_Os05g02500	GSN1/OsMKP1	153	LOC_Os08g39890	IPA1/OsSPL14
30	LOC_Os02g18370	TAP	92	LOC_Os05g06280	SRS3/BHS1	154	LOC_Os08g39890	IPA1/OsSPL14
31	LOC_Os02g27620	D50	93	LOC_Os05g06660	GS5	155	LOC_Os08g41940	GW8/SPL16
32	LOC_Os02g29040	ANK1	94	LOC_Os05g09520	qSW5/GW5	156	LOC_Os08g41950	OsMADS7/45
33	LOC_Os02g29160	ank	95	LOC_Os05g11730	GSK2	157	LOC_Os08g42540	OsOTUB1
34	LOC_Os02g29190	ank	96	LOC_Os05g26890	GA1/RGA1	158	LOC_Os09g13610	MED25/PFT1
35	LOC_Os02g29210	ANK2	97	LOC_Os05g31040	OsCKX9	159	LOC_Os09g13630	EMF2b
36	LOC_Os02g32950	RCN2	98	LOC_Os05g33730	gid1	160	LOC_Os09g26999	DEP1
37	LOC_Os02g34850	OsSET8	99	LOC_Os05g34854	Prl5/GA20ox2	161	LOC_Os09g28520	SG1
38	LOC_Os02g35180	RR2	100	LOC_Os05g38000	OsRopGEF10	162	LOC_Os09g31438	SPL17
39	LOC_Os02g36974	gid2	101	LOC_Os05g38120	VPB1/	163	LOC_Os09g32944	SPL18

KHOA HỌC & CÔNG NGHỆ

TT	Locus_id MSU	Tên gene	TT	Locus_id MSU	Tên gene	TT	Locus_id MSU	Tên gene
					SH5/RI			
40	LOC_Os02g41460	OsG1L3	102	LOC_Os05g38420	OsLAC15	164	LOC_Os09g32948	MADS24/ MADS8
41	LOC_Os02g45160	OsALMT 7	103	LOC_Os05g39880	OsEPFL8	165	LOC_Os09g34180	OsFH15
42	LOC_Os02g47280	GS2/ OsGRF4	104	LOC_Os05g41760	MSF1	166	LOC_Os09g36160	SHI1
43	LOC_Os02g50480	HK5	105	LOC_Os05g49590	SPA Small Panicle	167	LOC_Os09g37710	OsNPL4
44	LOC_Os02g51300	AP2/ERE BP22	106	LOC_Os05g51360	EBS	168	LOC_Os10g21810	HK3
45	LOC_Os02g51410	sk1	107	LOC_Os06g03710	OsGRAS- 32	169	LOC_Os10g28330	Ehd2/RID1/ OsId1
46	LOC_Os02g56630	GN2	108	LOC_Os06g03970	OsER2/ OsSIK1	170	LOC_Os10g32600	Ehd1/EF1
47	LOC_Os03g01890	LF1 / OsHB1	109	LOC_Os06g05060	Hd17	171	LOC_Os10g33780	TAW1
48	LOC_Os03g02160	EHD4	110	LOC_Os06g06050	D3	172	LOC_Os10g39130	MADS56
49	LOC_Os03g03070	DTH3/ OsSOC1	111	LOC_Os06g06300	RFT1	173	LOC_Os11g04190	OsPINc
50	LOC_Os03g05680	SE14	112	LOC_Os06g06530	HGW	174	LOC_Os11g05930	CCT38/ PRR59
51	LOC_Os03g06610	OsEPFL6	113	LOC_Os06g06750	OsMADS5	175	LOC_Os11g12740	SP1
52	LOC_Os03g07510	PGL1	114	LOC_Os06g10230	OsER1	176	LOC_Os11g14220	SRS5
53	LOC_Os03g12660	osdwarf4 -1	115	LOC_Os06g12610	OsPINa	177	LOC_Os11g29840	LA1
54	LOC_Os03g12950	PLT9	116	LOC_Os06g16370	Hd1	178	LOC_Os11g37190	OsEPFL7
55	LOC_Os03g17350	OsABCG 5	117	LOC_Os06g16390	SET24/ SDG711	179	LOC_Os11g37280	Osc6
56	LOC_Os03g19480	SDG718	118	LOC_Os06g19444	OsCCT22	180	LOC_Os11g38270	FON4/ TG1
57	LOC_Os03g20380	OsCIPK3 1	119	LOC_Os06g30370	MFT1	181	LOC_Os12g01120	NAL3
58	LOC_Os03g44500	qgl3/ GL3.1	120	LOC_Os06g35940	OsFTL12	182	LOC_Os12g04000	OsPINd
59	LOC_Os03g46930	OsEPFL4	121	LOC_Os06g37364	D35	183	LOC_Os12g16160	OsCCT41
60	LOC_Os03g49880	OsTB1/ FC1	122	LOC_Os06g39650	sped1	184	LOC_Os12g24080	OsUPL2
61	LOC_Os03g50860	OsK4 / pal1	123	LOC_Os06g39880	BRD3	185	LOC_Os12g31748	OsMADS20
62	LOC_Os03g51030	PHYA	124	LOC_Os06g40780	MOC1/ GNP6	186	LOC_Os12g34850	OsVIL2

2.2. Phương pháp nghiên cứu

2.2.1. Thu thập và phân tích dữ liệu SNP và INDEL

Bộ dữ liệu đầu vào của nghiên cứu là các tệp Variant Call Format (VCF) chứa thông tin biến dị SNP và INDEL của ba dòng lúa đột biến A64, A68, A83 và giống đối chứng D3. Các tệp VCF được xây dựng từ kết quả giải trình tự hệ gene và phân tích tin sinh học do Công ty TNHH Lobi Việt Nam thực hiện. Tệp VCF sử dụng trong nghiên cứu là tệp biến dị cuối cùng đã được đơn vị cung cấp dịch vụ xử lý, kiểm soát chất lượng và sàng lọc trước khi bàn giao. Trong nghiên cứu

này, đã sử dụng trực tiếp tệp VCF đã được xử lý và không thực hiện bổ sung các bước lọc biến dị hoặc thiết lập các ngưỡng lọc bằng VCFtools. Phần mềm VCFtools phiên bản 0.1.16 được sử dụng để chuẩn hóa dữ liệu và trích xuất các biến dị SNP và INDEL trên các gene mục tiêu phục vụ các phân tích tiếp theo. Sau đó, dữ liệu được xử lý trên phần mềm R phiên bản 4.3.2 (R Core Team, 2024) để đọc dữ liệu VCF, trích xuất thông tin về nhiễm sắc thể, vị trí biến dị, alen tham chiếu (REF), alen biến dị (ALT), kiểu biến dị (SNP hoặc INDEL) và kiểu gene của từng dòng vật liệu.

Dữ liệu sau xử lý được tổng hợp và thống kê theo từng dòng lúa nghiên cứu nhằm xác định số lượng SNP và INDEL của mỗi mẫu, đồng thời tạo cơ sở dữ liệu phục vụ các bước khai thác biến dị trên tập hợp các gene mục tiêu và lựa chọn các locus đa hình để phát triển chi thị phân tử.

2.2.2. Phân tích các gene ứng viên liên quan đến năng suất và chất lượng lúa

Danh sách 186 gene liên quan đến sinh trưởng, năng suất và chất lượng lúa được sử dụng làm tập gene mục tiêu để khai thác biến dị. Tọa độ của các gene được đối chiếu với thông tin vị trí biến dị trong các tệp VCF trên phần mềm R nhằm trích xuất toàn bộ các SNP và INDEL nằm trong vùng của từng gene.

Các biến dị sau khi trích xuất được phân loại thành SNP và INDEL, đồng thời thống kê theo từng gene và từng dòng vật liệu nghiên cứu. Số lượng biến dị của mỗi gene được tổng hợp và so sánh giữa các dòng A64, A68, A83 với giống đối chứng D3 nhằm xác định các gene có mức độ đa hình cao và các locus INDEL có tiềm năng ứng dụng trong phát triển chi thị phân tử.

Kết quả phân tích được sử dụng để đánh giá mức độ phân bố của SNP và INDEL trên các gene liên quan đến năng suất và chất lượng lúa, đồng thời làm cơ sở lựa chọn các biến dị INDEL phù hợp cho bước thiết kế chi thị phân tử.

2.2.3. Thiết kế chi thị INDEL ứng viên

Các biến dị INDEL được ưu tiên lựa chọn để thiết kế cặp mồi khi tạo ra sự sai khác kích thước sản phẩm PCR dự kiến từ 9 bp trở lên giữa các dòng vật liệu nghiên cứu. Ngưỡng này được lựa chọn nhằm tăng khả năng phân biệt sản phẩm PCR bằng điện di trên gel thông thường, đồng thời nâng cao khả năng nhận diện đa hình giữa

các dòng nghiên cứu. Trình tự ADN bao quanh vị trí biến dị INDEL được sử dụng để thiết kế cặp mồi bằng phần mềm ApE phiên bản 3.1.6. Độ đặc hiệu của các cặp mồi được kiểm tra bằng công cụ NCBI Primer-BLAST nhằm đảm bảo mỗi cặp mồi chỉ khuếch đại một vị trí mục tiêu trên hệ gene lúa.

Các cặp mồi được lựa chọn theo các tiêu chí: Chiều dài mồi từ 18-25 bp (tối ưu 20 bp), hàm lượng GC từ 40-60%, nhiệt độ nóng chảy (T_m) từ 57-63°C (tối ưu khoảng 60°C) và kích thước sản phẩm PCR dự kiến từ 100-400 bp. Đồng thời, các cặp mồi được đánh giá khả năng hình thành cấu trúc kẹp tóc (hairpin), hiện tượng tự bắt cặp (self-dimer) và bắt cặp chéo (hetero-dimer). Những cặp mồi đáp ứng các tiêu chí trên được lựa chọn để xây dựng bộ 42 chi thị INDEL ứng viên, phục vụ các nghiên cứu đánh giá đa hình và hỗ trợ chọn giống lúa.

2.3. Thời gian và địa điểm nghiên cứu

- Thời gian nghiên cứu: Năm 2025 - 2026.
- Địa điểm nghiên cứu: Viện Di truyền Nông nghiệp.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Phân tích biến dị SNP và INDEL trên các gene liên quan đến năng suất và chất lượng lúa

Từ bộ dữ liệu giải trình tự hệ gene của các dòng lúa đột biến A64, A68, A83 và giống đối chứng D3, tiến hành trích xuất các biến dị SNP và INDEL nằm trong vùng trình tự của 186 gene liên quan đến sinh trưởng, năng suất và chất lượng lúa. Kết quả phân tích xác định được tổng cộng 10.614 biến thể SNP và INDEL trong tập hợp các gene khảo sát, bao gồm: 3.479 biến thể ở dòng A64; 3.877 biến thể ở dòng A68 và 3.258 biến thể ở dòng A83 (Bảng 2).

Bảng 2. Tổng hợp số lượng biến dị SNP và INDEL ở các dòng lúa đột biến

Dòng vật liệu	Số lượng SNP/INDEL	Đặc điểm nổi bật
A64	3.479	Mức đa hình trung bình
A68	3.877	Số lượng biến dị cao nhất
A83	3.258	Đa hình trên nhiều gene ứng viên

Mức độ tích lũy biến dị có sự khác biệt đáng kể giữa các gene và giữa các dòng vật liệu nghiên cứu. Số lượng biến dị trên từng locus dao động từ vài biến thể đến trên 100 biến thể tùy theo vùng gene khảo sát. Phần lớn các gene ghi nhận mức đa hình trung bình, tuy nhiên một số locus xuất

hiện số lượng lớn SNP và INDEL, cho thấy đây có thể là các vùng hệ gene có mức độ đa hình cao ở các dòng đột biến. Đáng chú ý, dòng A68 ghi nhận tổng số biến dị cao nhất (3.877 biến thể), tiếp theo là A64 (3.479 biến thể) và A83 (3.258 biến thể). Kết quả này cho thấy, mức độ tích lũy

biến dị giữa các dòng đột biến không hoàn toàn giống nhau, phản ánh sự khác biệt về phổ biến dị di truyền của từng dòng vật liệu.

Kết quả cũng cho thấy, nhiều biến dị xuất hiện trên các gene liên quan trực tiếp đến cấu trúc bông, kiến trúc cây và kích thước hạt. Một số gene ghi nhận số lượng biến dị cao gồm: *ANK1*, *ANK2*, *GN2*, *AP01/SCM2*, *DEP1*, *IPA1/OsSPL14*, *GW2* và *GS5*. Đây là các gene đã được công bố có vai trò quan trọng trong điều khiển số nhánh bông, số hạt trên bông, kiểu khóm và kích thước hạt ở lúa [3, 4]. Việc phát hiện nhiều SNP và INDEL trên các gene này cho thấy, đây là nguồn biến dị di truyền có giá trị, có thể được khai thác để phát triển các chỉ thị phân tử và hỗ trợ nhận diện các allele liên quan đến năng suất và chất lượng lúa.

So sánh giữa các dòng đột biến và giống đối chứng D3 cho thấy, nhiều biến dị xuất hiện đặc trưng ở từng dòng vật liệu. Một số locus ghi nhận mức đa hình cao ở A68 và A83 nhưng thấp hơn ở A64, phản ánh sự khác biệt về mức độ tích lũy biến dị giữa các dòng đột biến. Sự khác biệt này cho thấy, mỗi dòng đột biến có thể mang các tổ hợp allele riêng biệt trên các gene liên quan đến năng suất và chất lượng lúa. Đây là cơ sở quan trọng để lựa chọn nguồn vật liệu phù hợp cho phát triển chỉ thị phân tử và khai thác nguồn gene trong các chương trình chọn giống.

Các nghiên cứu gần đây cho thấy, dữ liệu SNP và INDEL từ giải trình tự hệ gene có thể được sử dụng hiệu quả trong phát triển chỉ thị phân tử và hỗ trợ chọn giống cây trồng [5, 10]. Trong nghiên cứu này, các biến dị phát hiện được là nguồn dữ liệu quan trọng cho việc phát triển các chỉ thị phân tử ứng viên phục vụ đánh giá nguồn gene và chọn giống lúa bằng chỉ thị phân tử.

3.2. Phân tích các gene ứng viên liên quan đến năng suất và chất lượng lúa

Trong số 186 gene khảo sát, nhiều locus ghi nhận số lượng lớn biến dị SNP và INDEL ở các dòng lúa đột biến A64, A68 và A83. Đáng chú ý, nhiều gene đã được công bố có vai trò quan trọng trong điều khiển cấu trúc bông, số hạt trên bông, kiến trúc cây và kích thước hạt cũng xuất hiện mức đa hình cao, cho thấy tiềm năng khai thác các allele ứng viên phục vụ chọn giống lúa. Các gene này được lựa chọn để phân tích sâu do vừa

ghi nhận mức đa hình tương đối cao trong bộ dữ liệu nghiên cứu, vừa đã được chứng minh có vai trò quan trọng đối với các tính trạng nông sinh học chủ yếu của cây lúa. Các gene thuộc cụm ANK/ANK-TPR trên nhiễm sắc thể 2, bao gồm: *ANK1* (*LOC_Os02g29040*) và *ANK2* (*LOC_Os02g29210*), ghi nhận số lượng lớn biến dị ở cả ba dòng đột biến, đặc biệt ở A68 và A83. Các nghiên cứu trước đây cho thấy, các gene ANK/ANK-TPR có liên quan đến điều hòa cấu trúc bông và sự phát triển số nhánh bông ở lúa [13]. Sự xuất hiện của nhiều SNP và INDEL trên các locus này cho thấy, đây có thể là nguồn allele tiềm năng liên quan đến các tính trạng năng suất. Kết quả này cho thấy, cụm gene ANK/ANK-TPR là nguồn biến dị đáng chú ý và có thể được ưu tiên khai thác trong phát triển các chỉ thị phân tử phục vụ chọn giống.

Một số gene liên quan trực tiếp đến số hạt trên bông và năng suất cũng ghi nhận mức đa hình cao. Gene *GN2* và *AP01/SCM2* xuất hiện nhiều biến dị SNP và INDEL ở các dòng đột biến, đặc biệt ở A68. *AP01/SCM2* đã được xác định có vai trò trong điều khiển số lượng nhánh sơ cấp và năng suất ở lúa [14], trong khi *GN2* liên quan đến điều hòa số lượng hạt trên bông [15]. Việc đồng thời phát hiện nhiều SNP và INDEL trên các gene này cho thấy tiềm năng khai thác đa dạng các dạng biến dị di truyền phục vụ phát triển chỉ thị phân tử và nghiên cứu mối liên hệ giữa kiểu gene với kiểu hình.

Bên cạnh đó, nhiều gene liên quan đến kiến trúc cây và kích thước hạt cũng ghi nhận sự hiện diện của SNP và INDEL ở các dòng đột biến. Các gene *DEP1*, *IPA1/OsSPL14* và *MOC1/GNP6* xuất hiện nhiều biến dị ở một hoặc nhiều dòng vật liệu nghiên cứu. Đây là các gene đã được chứng minh có vai trò quan trọng trong điều hòa kiểu khóm, số nhánh hữu hiệu và cấu trúc tán cây ở lúa [16, 17]. Ngoài ra, các gene liên quan đến kích thước và hình dạng hạt như: *GW2*, *GS5*, *qSW5/GW5* và *GW8/SPL16* cũng ghi nhận mức đa hình tương đối cao. Những gene này đã được xác định có liên quan đến chiều dài, chiều rộng và khối lượng hạt ở lúa [18, 19]. Các kết quả này cho thấy, nguồn biến dị thu được không chỉ tập trung ở nhóm gene điều khiển năng suất mà còn phân bố trên các gene liên quan đến kiến trúc cây và chất lượng hạt, tạo nguồn vật liệu phong phú cho các nghiên cứu chọn giống bằng chỉ thị phân tử.

Bảng 3. Các gene ứng viên liên quan đến năng suất và chất lượng lúa

Nhóm tính trạng	Gene	Locus ID	Chức năng chính
Cấu trúc bông	<i>ANK1</i>	LOC_Os02g29040	Điều hòa số nhánh bông
Cấu trúc bông	<i>ANK2</i>	LOC_Os02g29210	Phát triển bông lúa
Số hạt trên bông	<i>GN2</i>	LOC_Os02g56630	Điều hòa số hạt
Năng suất	<i>APO1/SCM2</i>	LOC_Os06g45460	Điều hòa số nhánh sơ cấp
Kiến trúc cây	<i>DEP1</i>	LOC_Os09g26999	Điều hòa kiểu bông dày
Kiến trúc cây	<i>IPA1/OsSPL14</i>	LOC_Os08g39890	Kiến trúc cây lý tưởng
Kích thước hạt	<i>GW2</i>	LOC_Os02g14720	Điều hòa chiều rộng hạt
Kích thước hạt	<i>GS5</i>	LOC_Os05g06660	Điều hòa kích thước hạt
Kích thước hạt	<i>qSW5/GW5</i>	LOC_Os05g09520	Điều hòa chiều rộng hạt
Thời gian sinh trưởng	<i>Hd1</i>	LOC_Os06g16370	Điều hòa trổ bông

So sánh giữa các dòng đột biến cho thấy, mức độ tích lũy biến dị không hoàn toàn giống nhau tại cùng một locus. Một số gene xuất hiện số lượng biến dị cao ở A68 nhưng thấp hơn ở A64 và A83, phản ánh sự khác biệt về nguồn biến dị giữa các dòng vật liệu nghiên cứu. Kết quả này cho thấy, mỗi dòng đột biến có thể mang các tổ hợp allele khác nhau liên quan đến các tính trạng nông sinh học quan trọng.

Các nghiên cứu gần đây cho thấy, việc khai thác dữ liệu SNP và INDEL từ giải trình tự hệ gene có thể hỗ trợ hiệu quả cho phát triển chỉ thị phân tử và chọn giống cây trồng [5, 10]. Trong nghiên cứu này, các biến dị phát hiện được trên các gene ứng viên cung cấp nguồn dữ liệu có giá trị cho việc phát triển các chỉ thị phân tử ứng viên phục vụ nhận diện kiểu gene và hỗ trợ chọn giống

lúa bằng chỉ thị phân tử trong tương lai. Trên cơ sở đó, các locus INDEL phù hợp đã được lựa chọn để thiết kế bộ chỉ thị phân tử.

3.3. Phát triển các chỉ thị INDEL ứng viên phục vụ chọn giống phân tử

Từ tập hợp các biến dị SNP và INDEL đã xác định, nghiên cứu tiến hành lựa chọn các biến dị INDEL có sự khác biệt rõ rệt giữa các dòng đột biến và giống đối chứng D3 nhằm phát triển các chỉ thị phân tử ứng viên phục vụ nhận diện kiểu gene và hỗ trợ chọn giống lúa. Các biến dị được ưu tiên lựa chọn bao gồm: Các locus phân bố trên các gene liên quan trực tiếp đến cấu trúc bông, kiến trúc cây, số hạt trên bông và kích thước hạt (Bảng 4), có khả năng tạo đa hình giữa các dòng nghiên cứu và phù hợp để thiết kế cặp mồi PCR theo các tiêu chí đã trình bày ở mục 2.2.3.

Bảng 4. Các chỉ thị INDEL ứng viên phục vụ chọn giống lúa

Gene	Chỉ thị phân tử	Loại biến dị	Dòng đa hình	Tiềm năng ứng dụng
<i>qSW5/GW5</i>	CM130-5733445	INDEL	A64/D3	Chỉ thị phân tử kích thước hạt
<i>APO1/SCM2</i>	CM131-29125268	INDEL	A64-A68/D3	Chỉ thị phân tử năng suất
<i>Hd1</i>	CM131-10219549	INDEL	A68-A83/D3	Chỉ thị phân tử thời gian sinh trưởng
<i>DEP1</i>	CM134-14884130	INDEL	A83/D3	Chỉ thị phân tử kiến trúc cây
<i>D3</i>	CM131-3328473	INDEL	A64/D3	Chỉ thị phân tử kiểu khóm
<i>SPL10</i>	CM131-28737432	INDEL	A83/D3	Chỉ thị phân tử cấu trúc bông

Kết quả phân tích đã lựa chọn và thiết kế được 42 chỉ thị INDEL ứng viên trên các gene liên quan đến năng suất và chất lượng lúa (Bảng 5). Các chỉ thị được phát triển trên nhiều gene ứng viên quan trọng như: *GN2*, *APO1/SCM2*, *DEP1*, *Hd1*, *D3*, *qSW5/GW5*, *SPL10* và *Ghd8/DTH8*. Các cặp mồi được thiết kế nhằm khuếch đại các vùng trình tự

chứa biến dị INDEL có khả năng tạo ra sự khác biệt kích thước sản phẩm PCR giữa các dòng vật liệu nghiên cứu. Việc phát triển đồng thời nhiều chỉ thị trên các nhóm gene chức năng khác nhau góp phần mở rộng khả năng ứng dụng của bộ chỉ thị trong nhận diện kiểu gene và hỗ trợ chọn giống đối với nhiều tính trạng mục tiêu.

Kích thước sản phẩm PCR dự kiến giữa các dòng A64, A68, A83, D3 và trình tự *indica* tham chiếu cho thấy nhiều chỉ thị phân tử có khả năng phân biệt kiểu gene rõ rệt. Kích thước sản phẩm PCR dự kiến của các chỉ thị INDEL dao động với mức sai khác từ 9-36 bp giữa các dòng vật liệu nghiên cứu. Trong đó, các chỉ thị trên gene *Hdl*, *qSW5/GW5* và *DEP1* ghi nhận mức sai khác từ 15 bp trở lên, trong khi các chỉ thị trên *AP01/SCM2*, *D3* và *SPL10* tạo ra sự sai khác từ 9 - 12 bp nhưng vẫn đáp ứng tiêu chí lựa chọn để

phân biệt kiểu gene bằng điện di trên gel thông thường. Đặc biệt, các chỉ thị phân tử phát triển trên các gene *qSW5/GW5*, *AP01/SCM2*, *D3*, *SPL10*, *DEP1* và *Hdl* cho thấy khả năng đa hình tương đối cao giữa các dòng nghiên cứu (Bảng 4). Mức sai khác kích thước sản phẩm PCR dự kiến của các chỉ thị này tạo thuận lợi cho việc nhận diện kiểu gene bằng các phương pháp điện di thông dụng, góp phần giảm chi phí phân tích và nâng cao tiềm năng ứng dụng trong các chương trình chọn giống lúa.

Bảng 5. Danh sách các cặp môi INDEL ứng viên

TT	Chỉ thị	Forward primer (5'-3')	Reverse primer (5'-3')	A64 (bp)	A68 (bp)	A83 (bp)	D3 (bp)	WT (bp)	Thay đổi polypeptide	Gene
1	CM126-4788392	gcttcactatgttggtgac	ctactctaattggtagttcc	241	241	255	255	255	Không	OsGI
2	CM126-5670396	gcaagagacttcaatttcacg	tactgtagcatcacattgtc	202	202	202	202	192	Không	Gn1a
3	CM126-6961668	gacgactgatctgattcaga	atcgatctcgtggctgcaag	129	129	129	161	129	Không	OsFT-L1
4	CM126-8198497	gatatcaagtatctcgacgatg	gcaacatcgctaaccacag	119	119	119	119	128	Có	OsNP-L3
5	CM126-20132489	cactgaactgaaggagagag	gtggtgaaacaaagtacct	124	138	124	124	124	Không	dbsl
6	CM126-25584079	cagtgggtcaaagaccacgac	gtagtaagtgcagtaactggg	110	110	110	110	126	Không	LOG1
7	CM126-25584348	ctagcagcatgccacttg	gtgctgctcattttgttc	130	139	130	130	130	Không	
8	CM126-45789032	gaagctgacgaacttggtgg	gaacacaactactacttgcc	124	124	124	115	124	Có	SE13
9	CM126-45941694	ggttaggggtagtcattgc	cttccatgtctcacataggg	136	120	120	136	120	Không	RR4
10	CM127-17511786	gcatacatgaacaaagcagc	gggtgtgcaatactgatagc	173	164	164	173	173	Không	D50
11	CM128-33893765	ggggagtgaaagatgacaag	ctcactgccgtatgtcagtg	174	174	174	186	186	Không	OsH1
12	CM128-35210703	ttgttaggatacgtgttttca	gcgattgaaaattttcaagtc	182	204	204	204	204	Không	MAD-S14
13	CM129-412885	gggtcacaagtggacttag	ggataaaaggatatccatgct	127	104	127	104	127	Không	Nglf-1
14	CM129-14174776	catttgcttcactctctctc	gaataaacgataggagtcctag	195	195	183	195	195	Không	Awn-1
15	CM129-22862690	gatgaggcagcagcactac	gtgtgtccgcgtcagaac	187	163	163	163	163	Có	HBP1
16	CM130-5733445	gaactacatgtccaacacgc	ggccgtcttgaagtgaac	217	217	217	202	217	Có	qSW5
17	CM130-16517474	ggacatcatctcacacttc	gaagattaccgacaactccg	143	133	133	133	143	Không	GA1
18	CM130-24794298	gtgtgtgtgtgtgtgtgtg	gggggaaagggcatatccag	184	184	184	184	211	Không	OsEPF-L8
19	CM131-3328473	caccccaaatccctcaactg	ctccggcagggtccagtac	160	151	160	160	160	Có	D3
20	CM131-	cccgatcgatcaat	cttcgctttcagct	249	249	249	201	249	Không	D3

KHOA HỌC & CÔNG NGHỆ

TT	Chỉ thị	Forward primer (5'-3')	Reverse primer (5'-3')	A64 (bp)	A68 (bp)	A83 (bp)	D3 (bp)	WT (bp)	Thay đổi polypeptide	Gene
	3334710	tatcgtc	gggtg							
21	CM131-3592491	cacaccgatgtgcaaatga	ccacatcgagtaaa tccctc	321	355	355	321	321	Không	HGW
22	CM131-3775189	gaccgtttgattgaacgc	tcctatggaatgga caatcc	122	133	133	133	133	Không	OsMA DS5
23	CM131-10219549	caaggacgagga ggtggac	gttggtttcgatgcg gttgtc	169	169	202	202	202	Có	Hd1
24	CM131-10237782	gatgacgaccaac ctaccac	gtttgggtttctgg gcc	214	214	200	200	200	Không	SET24
25	CM131-10245622	ccgtgagaattcat cgttc	ggatgtaaacaact cgttc	133	153	153	153	153	Không	SET24
26	CM131-22731412	gatcaccatctgcc acatgac	tcgttcttgactctc gttagag	131	147	147	147	147	Không	D35
27	CM131-22735690	gtgtgatccagtga ttctcg	catttccattatctgc gaacag	143	125	125	125	143	Không	D35
28	CM131-22736373	ggtagtgagaga gatgcag	cttagtttgactcc cggc	169	169	169	169	148	Không	D35
29	CM131-28737432	catgatgagcggta ggatga	gaagtccaagccg tcgtac	181	181	193	181	193	Có	SPL10
30	CM131-29125268	gaactgcgcgtgc acctc	gaagagccgtctc agcgtg	115	124	124	115	124	Có	APO1/ SCM2
31	CM132-1392898	ccatttcatactcc atccatc	ggcccatacacac attcatc	181	181	203	203	203	Không	PLT8/ crl5
32	CM132-1614084	cacacatgtccttc tgcctc	catccaaagcaacc acctcc	158	146	158	158	158	Không	D6
33	CM133-4587937	ctacgcgtgtgcct agctag	cgagggtgatgga gttcagag	151	151	151	142	151	Có	Ghd8/ DTH8
34	CM133-26726628	gagcgaatcagga agctgc	cagatttgctgctac tactgc	157	141	141	157	141	Không	POH1
35	CM133-28251828	cttggaaccaattg ccggtc	cagggtacctgca aaacagg	151	170	170	170	170	Không	OsMA DS7
36	CM134-7117260	gtgacacatctat gtccagg	caagatcggtttgg aggttcc	135	135	122	135	122	Không	MED 25
37	CM134-14884130	gtgctttggaacca ctccct	ggtagttgctaag gccctg	182	182	182	167	167	Không	DEP1
38	CM134-14885258	ctagtcagggatg taatcatc	ctcagtttgccagta gtacag	122	122	122	122	143	Không	DEP1
39	CM135-19806999	ctaaccacatgacg acgacga	gccgggcaatcca atgtac	125	125	125	116	125	Có	MAD S56
40	CM137-86295	ccaccatcacccat acgctc	gatgcagtgtgca gtatgc	102	111	102	102	111	Có	NAL3
41	CM137-10796008	gcagctattgcaac attacttc	ctgatgcaactagc agcctag	179	158	158	158	158	Không	LARGE 2
42	CM137-10798446	gcgctttgctacgg gaattac	ctttgaggggagaac accgac	366	402	366	402	402	Không	LARGE 2

Các nghiên cứu trước đây cho thấy, chỉ thị INDEL là công cụ hiệu quả trong đánh giá đa dạng di truyền và chọn giống cây trồng do có tính ổn định cao, dễ phát hiện và chi phí phân tích thấp [8, 10]. Trong nghiên cứu này, các chỉ thị phân tử INDEL ứng viên được phát triển từ dữ liệu giải trình tự hệ gene có tiềm năng ứng dụng

trong nhận diện kiểu gen, đánh giá nguồn vật liệu và hỗ trợ chọn giống lúa bằng chỉ thị phân tử.

Mặc dù các chỉ thị phân tử trong nghiên cứu mới được đánh giá ở mức độ tin sinh học thông qua dữ liệu hệ gene và kích thước sản phẩm PCR dự kiến, kết quả thu được cho thấy tiềm năng lớn cho việc phát triển hệ thống chỉ thị phân tử phục

vụ nghiên cứu và chọn giống lúa trong tương lai. Các chỉ thị phân tử ứng viên này cần tiếp tục được kiểm chứng bằng thực nghiệm PCR và đánh giá trên tập đoàn vật liệu rộng hơn nhằm xác định tính ổn định và khả năng ứng dụng trong các chương trình chọn giống thực tiễn. Đây cũng là định hướng nghiên cứu tiếp theo nhằm hoàn thiện và đánh giá hiệu quả ứng dụng của bộ chỉ thị trong điều kiện sản xuất.

4. KẾT LUẬN VÀ KIẾN NGHỊ

4.1. Kết luận

Xác định được tổng cộng 10.614 biến thể SNP và INDEL phân bố trên nhiều gene liên quan đến cấu trúc bông, kiến trúc cây, số hạt trên bông và kích thước hạt.

Nhiều gene ứng viên quan trọng như: *ANK1*, *ANK2*, *GN2*, *AP01/SCM2*, *DEP1*, *IPA1/OsSPL14*, *GW2* và *GS5* ghi nhận mức đa hình cao giữa các dòng vật liệu nghiên cứu. Các biến dị phát hiện được không chỉ phản ánh sự đa dạng di truyền giữa các dòng lúa đột biến mà còn cung cấp nguồn dữ liệu có giá trị để khai thác các allele ứng viên liên quan đến năng suất và chất lượng lúa, làm cơ sở cho việc phát triển chỉ thị phân tử phục vụ nghiên cứu và chọn giống lúa.

Từ các biến dị INDEL đã thiết kế được 42 chỉ thị INDEL ứng viên trên các gene liên quan đến năng suất và chất lượng lúa. Nhiều chỉ thị phân tử cho thấy khả năng tạo ra sự khác biệt kích thước sản phẩm PCR giữa các dòng vật liệu nghiên cứu, có tiềm năng ứng dụng trong nhận diện kiểu gen, đánh giá nguồn vật liệu và hỗ trợ chọn giống lúa bằng chỉ thị phân tử.

4.2. Kiến nghị

Bổ sung nguồn dữ liệu biến dị di truyền trên các dòng lúa đột biến triển vọng và tạo cơ sở cho việc phát triển các chỉ thị phân tử phục vụ đánh giá nguồn gene và chọn giống lúa trong tương lai.

Các chỉ thị phân tử ứng viên cần tiếp tục được kiểm chứng bằng thực nghiệm và đánh giá trên tập đoàn vật liệu rộng hơn nhằm xác định tính ổn định và khả năng ứng dụng trong các chương trình chọn giống thực tiễn.

LỜI CẢM ƠN

Nghiên cứu được thực hiện với sự hỗ trợ kinh phí của nhiệm vụ: “Ứng dụng công nghệ sinh học

và năng lượng bức xạ trong chọn tạo giống lúa mới năng suất cao, chất lượng tốt phục vụ nội tiêu và xuất khẩu”, mã số: 20230923006 thuộc Đề án phát triển công nghiệp sinh học ngành nông nghiệp đến năm 2030, Bộ Nông nghiệp và Môi trường. Nhóm tác giả xin trân trọng cảm ơn.

TÀI LIỆU THAM KHẢO

- [1]. Yang X, Yu S, Yan S, Wang H, Fang W, Chen Y, Ma X, Han L (2024). Progress in rice breeding based on genomic research. *Genes*, 15(5), 564.
- [2]. Ray DK, Mueller ND, West PC, Foley JA (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS ONE*, 8(6), e66428.
- [3]. Wang C, Han B (2022). Twenty years of rice genomics research: From sequencing and functional genomics to quantitative genomics. *Molecular Plant*, 15(4), 593-619.
- [4]. Huang X, Han B (2014). Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology*, 65, 531-551.
- [5]. Kumar R, Prasad S. D., Burhan U. C. (2024). Advances in genomic tools for plant breeding: Harnessing DNA molecular markers, genomic selection, and genome editing. *Biological Research*, 57, 80.
- [6]. Hasan N, Choudhary S, Naaz N, Sharma N, Laskar RA (2021). Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *Journal of Genetic Engineering and Biotechnology*, 19, 1-18.
- [7]. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, 20(8), 467-484.
- [8]. Li P, Li Z., Liu X., Zhang H., Liu F., Li N., Yang Y., Yao F. (2023). Haplotype analysis and marker development of five salt-tolerant related genes in rice. *Frontiers in Plant Science*, 14, 1259462.
- [9]. Varshney RK, Terauchi R, McCouch SR (2014). Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. *PLoS Biology*, 12, e1001883.
- [10]. Pariasca-Tanaka J, Ueda, Y., Kondo K. (2025). Genome-wide sequence comparison and

development of InDel and SNP markers to facilitate localized rice breeding. *Current Plant Biology*, 42, 100469.

[11]. Tam N. T., Nhan D. K. (2024). Identification of insertion/deletion markers for photoperiod sensitivity in rice (*Oryza sativa* L.). *Biology*, 13(5), 358.

[12]. Sikora P., Chawade A., Larsson M., Olsson J., Olsson O. (2011). Mutagenesis as a tool in plant genetics, functional genomics, and breeding. *International Journal of Plant Genomics*, Article ID 314829, 13. DOI:10.1155/2011/314829

[13]. Lyu J. (2020). Neo-functionalization of a domestication gene confers key morphological differences between indica and japonica rice. *Nature Communications*, 11, 5474.

[14]. Ookawa T., Hobo T., Yano M., Murata K. (2010). New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield. *Nature Communications*, 1, 132.

[15]. 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*, 309, 741-745.

[16]. Jiao Y., Wang Y., Xue D., Wang J., Yan M., Liu G., Dong G., Zeng D., Lu Z., Zhu X., Qian Q., Li J. (2010). Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nature Genetics*, 42, 541-544.

[17]. 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.

[18]. Song XJ., Huang W., Shi M., Zhu MZ., Lin HX (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nature Genetics*, 39, 623-630.

[19]. 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.

EXPLOITATION OF SNP AND INDEL VARIATIONS FOR MOLECULAR MARKER DEVELOPMENT IN PROMISING RICE MUTANT LINES

Khong Ngan Giang¹, Vo Thi Minh Tuyen¹, Nguyen Thi Hao¹, Doan Van Son¹,
Tran Thi Hue Huong², Nguyen Truong Giang³, Tran Manh Cuong³, Le Duc Thao¹

¹Agricultural Genetics Institute

²Vietnam Academy of Agricultural Sciences

³Ministry of Agriculture and Environment

Abstract

This study was conducted to exploit SNP and INDEL variations for molecular marker development in promising rice mutant lines. Whole-genome sequencing data from three rice mutant lines, A64, A68 and A83, together with the control variety D3, were used to analyze genetic variations in 186 genes associated with growth, yield and grain quality traits in rice. SNP and INDEL variants located within gene regions were extracted, filtered and compared among mutant lines and the control variety to identify potential polymorphic loci for molecular marker development. A total of 10,614 SNP and INDEL variants were identified in the investigated gene set, including 3,479 variants in A64, 3,877 variants in A68 and 3,258 variants in A83. Several genes related to panicle architecture, plant architecture and grain size showed high levels of polymorphism, including *ANK1*, *ANK2*, *GN2*, *APO1/SCM2*, *DEP1*, *IPA1/OsSPL14*, *GW2* and *GS5*. Based on the identified variants, 42 candidate INDEL primer pairs were designed for genotype discrimination and molecular marker development. The results provide valuable genetic variation resources and potential molecular markers for germplasm evaluation and marker-assisted rice breeding in future studies.

Keywords: Rice mutant lines, SNP, INDEL, molecular markers, rice breeding, grain yield.

Ngày nhận bài: 15/5/2026

Ngày chuyển phản biện: 25/5/2026

Ngày thông qua phản biện: 23/6/2026

Ngày duyệt đăng: 6/7/2026