ỨNG DỤNG CÔNG NGHỆ GENE CAPTURE ĐỂ XÁC ĐỊNH CÁC BIẾN THỂ DI TRUYỀN TRONG VÙNG QTL9 LIÊN QUAN ĐẾN CẤU TRÚC BÔNG LÚA

Stefan Juannic[1], Phạm Thị Mai[2], Lê Thị Như[2], Phạm Xuân Hội[2], Khổng Ngân Giang[2*]

[1] Đại học Montpellier, Đơn vị nghiên cứu DIADE, Viện Nghiên cứu phát triển (IRD), Montpellier, Pháp
[2] Phòng Thí nghiệm trọng điểm Công nghệ Tế bào thực vật, Viện Di truyền Nông nghiệp, Viện Khoa học Nông nghiệp Việt Nam

TÓM TẮT

QTL9 là một QTL tiềm năng mới liên quan đến cấu trúc bông lúa, chọn lọc được từ kết quả phân tích GWAS các tính trạng năng suất của tập đoàn lúa bản địa Việt Nam. Hai giống lúa G6 và G189, nằm trong tập đoàn nghiên cứu GWAS và thuộc hai haplotype khác biệt, được sử dụng làm bố mẹ để tạo quần thể lai tái tổ hợp nhằm nghiên cứu chức năng của QTL9. Trong nghiên cứu này, chúng tôi ứng dụng công nghệ Gene Capture kết hợp với giải trình tự thế hệ mới (Illumina) để tìm kiếm các biến thể di truyền trong vùng QTL9 của hai giống lúa G6 và G189. Kết quả phân tích và sàng lọc biến thể di truyền đã thu được 1002 biến thể đồng hợp trong đó có 827 biến thể SNPs, 175 biến thể INDELs nằm trên các vùng exon, intron, promoter, UTR-3', UTR-5' và vùng giao thoa giữa các gen. Từ đó, xác định được 12 SNPs nằm trong vị trí cắt của 5 enzyme giới hạn (*Sac*I, *Dra*I, *EcoR*V, *BamH*I, *Sal*I) để phát triển chỉ thị phân tử CAPS (Cleaved Amplified Polymorphic Sequences), phục vụ các nghiên cứu chọn tạo giống lúa năng suất cao ở Việt Nam.

Từ khóa: Cấu trúc bông lúa, Công nghệ Gene Capture, giải trình tự thế hệ mới NGS, QTL9, SNP.

MỞ ĐẦU

Cây lúa (*Oryza sativa.* L) cung cấp lương thực cho hơn một nửa dân số thế giới (Gross *et al*., 2014) và cũng là cây mô hình trong các nghiên cứu hệ gen cho các loài một lá mầm khác. Hơn 3000 giống lúa đã được giải trình tự, bao gồm Nipponbare (Kawahara *et al*., 2013), 93-11 (Yu *et al*., 2002), DJ 123, IR64 (Schatz *et al*., 2014), Zhenshan 97, Minghui 63 (Zhang *et al*., 2016), Shuhui 498 (Du *et al*., 2017), *Oryza glaberrima* (Wang *et al*., 2014; Wang *et al*., 2018), hai giống lúa thơm thuộc nhóm Japonica của Việt Nam là Tám Xoan Hải Hậu và Tám Xoan Bắc Ninh cũng đã được giải trình tự (Trung *et al*., 2017). Sự sẵn có của các bộ dữ liệu giải trình tự tạo cơ sở thuận lợi cho các nghiên cứu trên toàn hệ gen và chọn tạo giống lúa (Jain *et al*., 2019).

Những tiến bộ trong công nghệ giải trình tự thế hệ thứ hai và kỹ thuật tin sinh học trong thập kỷ qua là cơ sở để phát triển các phương pháp nghiên cứu đánh giá đa dạng di truyền cho nhiều loài thực vật. Tuy nhiên, đối với những cây có bộ gen lớn và có tính lặp lại cao bao gồm nhiều loại cây ngũ cốc, các thách thức kỹ thuật và chi phí có thể tạo thành rào cản cho việc giải trình tự cả bộ genome với độ phân giải cao, nhất là trên quy mô quần thể. Các phương pháp giải trình tự "rút gọn" chỉ tập trung vào một vùng cụ thể trong genome, bao gồm giải trình tự exome, giải trình tự RNA (RNA-seq) và giải trình tự vùng mục tiêu làm giảm đáng kể lượng dữ liệu tạo ra cũng như công việc xử lý tin sinh, có thể được áp dụng cho bất kỳ loài nào đã có trình tự của giống tham chiếu (Kawahara et al. 2012). Chụp gen kết hợp với giải trình tự thế hệ mới, là phương pháp hiệu quả để khám phá các vùng di truyền mục tiêu ở độ phân giải cao và qua đó xác định nhanh chóng hàng ngàn đa hình di truyền (Hill *et al*., 2019).

Trong nghiên cứu này, chúng tôi sử dụng phương pháp chụp gen kết hợp với giải trình tự thế hệ mới (Illumina) để xác định các biến thể di truyền (SNP - Single Nucleotide Polymorphism, INDELs) trong vùng QTL9 (Quantitive Trait Loci 9) liên quan đến cấu trúc bông lúa. QTL9 nằm trên nhiễm sắc thể số 2, có chiều dài 780 kb, được chọn lọc thông qua phân tích liên kết trên toàn hệ gen (GWAS Geneome Wide Association Study) của tập đoàn lúa bản địa Việt Nam (Ta *et al*., 2018). QTL9 liên kết với cả 2 tính trạng số hạt/bông và số gié thứ cấp/bông, là hai tính trạng quan trọng quyết định đến năng suất lúa. Đây là một QTL hoàn toàn mới, hơn 130 gen tìm thấy trong vùng QTL9 nhưng chưa gen nào được nghiên cứu chức năng liên quan đến cấu trúc bông. Tuy nhiên kết quả phân tích GWAS chủ yếu dựa vào các thuật toán thống kê, để có thể ứng dụng vào các chương trình lai cải tạo năng suất, QTL9 cần được nghiên cứu chức năng và phát triển các chỉ thị liên kết thông qua các quần thể lai. Hai giống lúa bản địa G6 (Sớm Giai Hưng yên) và G189 (Khẩu Nam Rinh) thuộc hai haplotype khác biệt và có cấu trúc bông tương phản (bông to và bông nhỏ) được sử dụng làm bố mẹ để tạo quần thể lai (Vũ Thị Nhiên *et al*. 2018). Kết quả phân tích biến thể di truyền trong vùng QTL9 của 2 giống lúa bố mẹ đã thu được 1002 biến thể đồng hợp. Trong đó có 12 SNPs nằm trong vị trí cắt của 5 enzyme giới hạn (*Sac*I, *Dra*I, *EcoR*V, *BamH*I, *Sal*I) được chọn lọc ra nhằm mục đích phát triển chỉ thị phân tử CAPS (Cleaved Amplified Polymorphic Sequences), phục vụ việc đánh giá sự phân ly của QTL9 trong quần thể F2 và các nghiên cứu chọn tạo giống lúa năng suất cao.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Nguyên liệu

Hai giống lúa G6 và G189 thuộc tập đoàn lúa bản địa Việt Nam sử dụng để nghiên cứu GWAS và được xác định thuộc 2 haplotype khác biệt về QTL9 (Ta *et al*., 2018). Giống G6 chứa 9 SNP (gagagcgaa) có kiểu hình bông nhỏ, số gié thứ cấp/bông dao động từ 23 - 34 gié, số hạt/bông từ 130 - 187 hạt; giống G189 chứa 9 SNP (atataaatt) có kiểu hình bông to, số gié thứ cấp/bông khoảng 45 - 48 gié và số hạt/bông từ 220 - 250 hạt (Vũ Thị Nhiên *et al*. 2018).

Trình tự gen giống lúa tham chiếu Nipponbare (Kawahara *et al*., 2013; (Matsumoto *et al*. 2016).

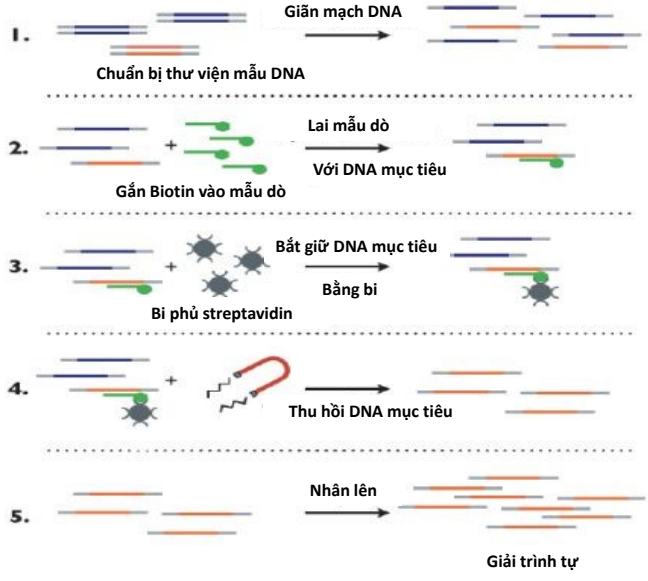
Phương pháp

Tách chiết DNA: DNA tổng số được tách chiết từ lá của các cây lúa non 2 tuần tuổi, sử dụng bộ kit DNeasy Plant Mini Kit (Qiagen) theo hướng dẫn của nhà sản xuất. Nồng độ và độ tinh sạch của DNA được kiểm tra trên máy Nanodrop. Chất lượng DNA được kiểm tra trên gel agarose 1% (100 V, TBE 0,5X, 60 phút).

Xây dựng thư viện mẫu dò RNA biotin hóa: Vùng QTL9 chứa 137 gen, mẫu dò được thiết kế bao phủ 77 gen mã hóa, kể cả vùng promoter (1000 pb trước mã mở đầu) và vùng 3' -UTR (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 oligonucleotide. Các sợi đơn DNA sau đó được phiên mã "*in vitro*" thành RNA và được biotin hóa.

Xây dựng thư viện DNA mục tiêu và làm giàu vùng mục tiêu: 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 điện di Fragment Analyser. DNA được tinh sạch bằng các hạt từ tính Agencourt AMPure XP SPRI (Beckman Coulter, Australia), 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.

Quá trình lai mẫu dò và 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") (Hình 1)**:** (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 1. Quá trình lai mẫu dò và vùng mục tiêu (theo MyBaits V2 của MYcroarray)

Giải trình tự: Các phân đoạn DNA phân lập được trên máy giải trình tự thế hệ mới Illumina Miseq en pair end (2 x 250).

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 gióng hàng với bộ gen tham chiếu của giống Nipponbarre (Kawahara *et al*., 2013) và loại bỏ vị trí phân tử trùng lặp, sử dụng phần mềm BWA mem. Kết quả gióng hàng dữ liệu sẽ được sàng lọc bằng phần mềm SAMtools để đảm bảo các đoạn trình tự đọc đã được gióng hàng chính xác với bộ gen tham chiếu và loại bỏ những sai xót. Các biến thể (SNPs, indels) sẽ được phát hiện bằng công cụ Haplotypecaller.

Gọi và sàng lọc biến thể*:* Gọi các biến thể bằng công cụ TOGGLe (Monat *et al*., 2015). Quá trình lọc để lựa chọn SNP đồng hợp và loại bỏ SNP dị hợp và trùng lặp được thực hiện bằng cách sử dụng VCFtools 0.1.16 (Danecek *et al*., 2011), SAMtools 1.9 (Li *et al*., 2009) và GATK 4.0.0.0 (DePristo *et al*., 2011; Auwera *et al*., 2013) theo các tiêu chí: Số lần đọc tối thiểu để gắn 1 biến thể dị hợp vào một giống (-minalcov = 3); Độ bao phủ tối thiểu (5 lần đọc) và độ bao phủ 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 (mising = 3). Các SNPs chất lượng và biallelic được giữ lại cho các phân tích tiếp theo.

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 plateforme Southgreen, sử dụng SNiPlay3, một ứng dụng trực tuyến để khám phá và phân tích quy mô lớn về các biến thể gen (Dereeper *et al*., 2011).

KẾT QUẢ VÀ THẢO LUẬN

Phân tích trình tự dựa trên bộ gen tham chiếu Nipponbare

Tổng cộng có 1.830.925 trình tự đọc đã được tạo ra. 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, adaptator và index được loại bỏ bằng công cụ Cutadapt. Từ hơn 1,8 triệu lượt đọc trình tự được truy xuất, gióng hàng dữ liệu với bộ gen tham chiếu Nipponbare và loại bỏ những phân tử trùng lặp, thu được hơn 98% số đoạn trình tự được giữ lại, trong đó 68,67-71% dữ liệu được ánh xạ vào vùng mục tiêu QTL9 (Bảng 1).

Bảng 1. So sánh và phân tích trình tự

Tên mẫu	Số đoạn trình tự gióng hàng thành công	Số đoạn trình tự gióng 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%)	525.416 (68,67%)

Xác định, sàng lọc và chú giải biến thể

Các biến thể (SNPs, indels) được phát hiện bằng công cụ Haplotypecaller. Số biến thể thu được lần lượt là 4524 và 4522 cho 2 giống G6 và G189. Sau đó những biến thể có tần suất allen 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 2 giống. So sánh với bộ dữ liệu di truyền DArt (Diversity Arrays Technology) (Phung *et al*., 2014), bộ dữ liệu GBS (Genotyping By Sequencing) (Phung *et al*., 2016) và dữ liệu kiểu hình cấu trúc bông của 2 haplotype (Ta *et al*., 2018), những biến thể không có mặt trong các bộ dữ liệu này bị loại bỏ. Bộ dữ liệu cuối cùng thu được giữa 2 haplotype có 1002 biến thể đồng hợp trong đó có 827 biến thể SNPs và 175 biến thể INDELs (Bảng 2).

Bảng 2. Kết quả gọi và sàng lọc biến thể

Tên biến thể	Tổng số	Vị trí					
		Exon	Intron	Promoter	UTR-3’	UTR-5’	Vùng giao thoa giữa các gen
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

Các biến thể trong vùng exon (117 biến thể) được 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... 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 plateforme Southgreen, sử dụng SNiPlay3, một ứng dụng trực tuyến để khám phá và phân tích quy mô lớn về các biến thể gen (Dereeper *et al*., 2011).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 (Bảng 3).

Bảng 3. Phân loại biến thể

Hiệu ứng \ Tên biến thể	SNP	INDEL
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

Sàng lọc SNPs nằm trong vị trí cắt của enzyme giới hạn để xây dựng chỉ thị CAPS

Phần mềm CAPSDETECTOR được sử dụng để sàng lọc các SNPs nằm trong vị trí cắt của enzyme giới hạn để xây dựng chỉ thị phân tử CAPS, phục vụ việc đánh giá kiểu gen của quần thể F2. Kết quả thu được 12 SNP nằm trong vị trí cắt của 5 enzyme giới hạn (*Sac*I, *Dra*I, *EcoR*V, *BamH*I, *Sal*I) (Bảng 4). Trong đó, có 6 SNP nằm trong vị trí cắt của enzyme giới hạn *Dra*I, 2 SNP nằm trong vị trí cắt của enzyme *BamH*I, 2 SNP nằm trong vị trí cắt của enzyme *EcoR*V, 1 SNP nằm trong vị trí cắt của enzyme *Sac*I và 1 SNP nằm trong vị trí cắt của enzyme *Sal*I. Các SNP này chủ yếu nằm trong vùng promoter (7 SNP), chỉ có 2 SNP nằm trong vùng intron, 1 SNP nằm trong vùng 3'-UTR và 1 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).

Bảng 4. Danh sách SNP nằm trong vị trí cắt của các enzyme giới hạn

STT	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	*Sac*I
2	16805351	LOC_Os02g28410	Promoter			GG	TT	*Dra*I
3	16882053	LOC_Os02g28530	Intron			AA	TT	*Dra*I
4	16919772	intergenic				TT	TATA	*Dra*I
5	16961918	LOC_Os02g28670	Promoter			AA	TT	*EcoR*V
6	17035526	LOC_Os02g28810	Promoter			AA	GG	*BamH*I
7	17039753	LOC_Os02g28820	Promoter			TT	CC	*EcoR*V
8	17124682	LOC_Os02g28910	Promoter			TATA	TAT	*Dra*I
9	17161487	LOC_Os02g28980	Promoter			TATA	TT	*Dra*I
10	17205540	LOC_Os02g29040	Promoter			AA	TT	*Sal*I
11	17254246	LOC_Os02g29130	intron			TT	AA	*Dra*I
12	17271258	LOC_Os02g29150	3'-UTR			GG	CC	*BamH*I

Ghi chú chữ viết tắt: H1: Haplotype 1, H2: Haplotype 2; P: Proline, L: Leucine.

KẾT LUẬN

Từ hơn 1,8 triệu lượt đọc trình tự được truy xuất, gióng hàng dữ liệu với bộ gen tham chiếu Nipponbare và loại bỏ những phân tử trùng lặp, thu được hơn 98% số đoạn trình tự được giữ lại, trong đó 68,67 - 71% dữ liệu được ánh xạ vào vùng QTL9.

Gọi biến thể bằng công cụ Haplotypecaller thu được 9.046 biến thể trong vùng QTL9, trong đó ở giống lúa G6 có 4524 biến thể và ở giống lúa G189 có 4522 biến thể. Sàng lọc biến thể thu được 1002 biến thể đồng hợp của 2 haplotype trong đó có 827 biến thể SNP và 175 biến thể INDELs. Trong 827 biến thể SNP có 107 biến thể nằm trong vùng exon, 325 biến thể nằm trong vùng intron, 256 biến thể nằm trong vùng promoter, 38 biến thể nằm trong vùng UTR-3' và 6 biến thể nằm trong vùng UTR-5', còn lại 95 biến thể nằm trong vùng giao thoa giữa các gen.

Sàng lọc các biến thể nằm trong vị trí cắt của enzyme giới hạn thu được 12 SNPs nằm trong vị trí cắt của 5 enzyme giới hạn (*Sac*I, *Dra*I, *EcoR*V, *BamH*I, *Sal*I) phục vụ cho việc phát triển chỉ thị phân tử CAPS nhằm phân tích sự phân ly của QTL9 trong quần thể F2.

Lời cảm ơn: *Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Quốc gia (NAFOSTED) thông qua đề tài mã số 106-NN.02-2016.60. Các tác giả xin chân thành cảm ơn.*

TÀI LIỆU THAM KHẢO

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, *et al.* (2011). The variant call format and VCFtools. *Bioinfor* 27(15): 2156-58.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5): 491-98.

Dereeper A, Nicolas S, Le Cunff L, Bacilieri R, Doligez A, *et al.* (2011). SNiPlay: A web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. *BMC bioinformatics* 12: 134.

Du H, Yu Y, Ma Y, Gao Q, Cao Y, *et al.* (2017). Sequencing and de novo assembly of a near complete indica rice genome. *Nature Commun* 8.

Gross BL, Zhao Z (2014). Archaeological and genetic insights into the origins of domesticated rice. *Proc Natl Acad Sci USA* 111(17): 6190-97

Hill CB, Wong D, Tibbits J, Forrest K, Hayden M, *et al.* (2019). Targeted enrichment by solution-based hybrid capture to identify genetic sequence variants in barley. *Sci Data* 6(1): 1-8.

Jain R, Jenkins J, Shu S, Chern M, Martin JA, *et al.* (2019). Genome sequence of the model rice variety KitaakeX. *BMC Genom* 20.

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, *et al.* (2013). Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice (NY)* 6.

Kawahara Y, Oono Y, Kanamori H, Matsumoto T, Itoh T, Minami E. 2012. Simultaneous RNA-Seq Analysis of a Mixed Transcriptome of Rice and Blast Fungus Interaction. *PLoS One* 7(11):

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinfor* 25(16): 2078-79.

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 (NY)* 9.

Monat C, Tranchant-Dubreuil C, Kougbeadjo A, Farcy C, Ortega-Abboud E, *et al.* 2015. TOGGLE: toolbox for generic NGS analyses. *BMC Bioinfor* 16: 374.

myBaits - Hyb Capture Kits. Arbor Biosciences. https://arborbiosci.com

Phung NTP, Mai CD, Hoang GT, Truong HTM, Lavarenne J, *et al.* 2016. Genome-wide association mapping for root traits in a panel of rice accessions from Vietnam. *BMC Plant Biol* 16(1): 64.

Phung NTP, Mai CD, Mournet P, Frouin J, Droc G, *et al.* (2014). Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. *BMC Plant Biol* 14(1): 371.

Schatz MC, Maron LG, Stein JC, Wences AH, Gurtowski J, *et al.* (2014). Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biol*. 15(11):

Ta KN, Khong NG, Ha TL, Nguyen DT, Mai DC, *et al.* (2018). A genome-wide association study using a Vietnamese landrace panel of rice (Oryza sativa) reveals new QTLs controlling panicle morphological traits. *BMC Plant Biol* 18(1): 282.

Trung KH, Nguyen TK, Khuat HBT, Nguyen TD, Khanh TD, *et al.* (2017). Whole Genome Sequencing Reveals the Islands of Novel Polymorphisms in Two Native Aromatic Japonica Rice Landraces from Vietnam. *Genome Biol Evol* 9(6): 1816-20.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, *et al.* (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinfor* 11(1110): 11.10.1-11.10.33.

Vũ TN, Tạ KN, Stefan J, Lê HL, Phạm XH, *et al.* (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* 11(96):

Wang M, Yu Y, Haberer G, Marri PR, Fan C, *et al.* (2014). The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat Genet* 46(9): 982-88.

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, *et al.* (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557(7703): 43-49.

Yu J, Hu S, Wang J, Wong GKS, Li S, *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Sci* 296(5565): 79-92.

Zhang J, Chen LL, Xing F, Kudrna DA, Yao W, *et al.* (2016). Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proc Natl Acad Sci USA* 113(35): E5163-71.

APPLICATION OF GENE CAPTURE TECHNOLOGY TO IDENTIFY GENETIC VARIANTS IN QTL9 RELATED TO RICE PANICLE STRUCTURE

Stefan Jouannic[1], Thi Mai Pham[2], Thi Nhu Le[2], Xuan Hoi Pham[2], Ngan Giang Khong[2*]

[1] University of Montpellier, UMR DIADE, IRD, Montpellier, France

[2] National Key Laboratory of Plant Cell Biotechnology, Agricultural Genetics Institute, Vietnam Academy of Agricultural Sciences

SUMMARY

QTL9 is a new potential QTL related to rice panicle structure, selected from a GWAS analysis of yield traits using a Vietnamese local rice panel. In order to validate the GWAS site, a bi-parental population was developed using the low branching Sớm Giai Hưng Yên (G6) and the high branching Khẩu Nam Rinh (G189) accessions from 2 haplotypes characterized by a contrasted phenotype for the spikelet number and sencondary branch number traits. In this study, Gene Capture technology combined with next generation sequencing (Illumina) was applied to identify genetic variations in the QTL9 region of G6 and G189 varieties. Analysis and screening of genetic variants yielded 1002 homologous variants, including 827 SNPs and 175 INDELs. Of which, 12 SNPs are located in the cleavage sites of 5 restriction enzymes (*Sac*I, *Dra*I, *EcoR*V, *BamH*I, *Sal*I) serving to develop CAPS (Cleaved Amplified Polymorphic Sequences) markers for high yield rice breeding programs in Vietnam.

Keywords: Rice panicle structure, Gene Capture technology, Next Generation Sequencing, QTL9, SNP.

[*] Author for correspondence: Tel: +84-86.8300769; Email: ngangiang.khong2010@gmail.com