

DOI:10.22144/ctu.jvn.2021.124

CÁC BIẾN THỂ GENE *OsTZF1* LIÊN QUAN ĐẾN KHẢ NĂNG CHỊU MẶN Ở GIỐNG LÚA ĐỐC PHỤNG BẰNG PHƯƠNG PHÁP GIẢI TRÌNH TỰ BỘ GENE

Huỳnh Kỳ^{1*}, Trần Đăng Thành Phát¹, Nguyễn Thị Kim Phụng¹, Văn Quốc Giang¹, Nguyễn Văn Mạnh¹, Trần In Đô¹, Nguyễn Thành Tâm², Nguyễn Châu Thanh Tùng¹, Nguyễn Lộc Hiền¹ và Huỳnh Như Điền¹

¹Khoa Nông nghiệp, Trường Đại học Cần Thơ

²Viện Nghiên cứu Phát triển Đồng bằng sông Cửu Long, Trường Đại học Cần Thơ

*Người chịu trách nhiệm về bài viết: Huỳnh Kỳ (email: hky@ctu.edu.vn)

Thông tin chung:

Ngày nhận bài: 06/02/2021

Ngày nhận bài sửa: 12/04/2021

Ngày duyệt đăng: 20/08/2021

Title:

Variation of *OsTZF1* gene related to salt tolerance in Doc Phung rice variety using whole genome sequencing

Từ khóa:

Đốc Phụng, giải trình tự bộ gene, InDel, SNP

Keywords:

Doc Phung, InDel, SNP and Whole genome sequencing

ABSTRACT

In this study, the next generation sequencing technology was used to resequence the genome of Doc Phung rice varieties (salt-tolerant variety) and Nep Mo (salt-susceptible variety) to identify functional markers that are involved in salt tolerance mechanisms in Doc Phung rice variety. In comparison with the reference genome, the result showed that Doc Phung genome was consisted of 1,918,726 variations of SNP and 163,409 InDels (81,435 insertions, and 81,974 deletion). Whereas in Nep Mo variety, there were 1,931,380 SNPs and 171,663 InDels (88,473 insertions and 83,190 deletion). Most of the variants are located in non-functional regions including upstreams, downstream, and intergenic, accounting for over 75%. The variation of *OsTZF1* (LOC_Os05g10670.1) gene that regulates the expression of those gene related to biological and abiotic stress factors, showed that there were 7 SNPs and 9 nucleotides insertion (encode 3 amino acid arginine) in Doc Phung variety when being compared to Nep Mo based on reference genome. This information will help the breeders to apply as a molecular marker, using salt-tolerant rice breeding program in the future.

TÓM TẮT

Trong nghiên cứu này, kỹ thuật giải trình tự thế hệ mới (next generation sequencing) được ứng dụng để giải trình tự của bộ gene 2 giống lúa Đốc Phụng (giống chống chịu mặn) và giống Nép Mỡ (giống mẫn cảm với mặn), nhằm tìm các chỉ thị phân tử là gene chức năng mà các gene này liên quan đến cơ chế chống chịu mặn có trong giống lúa Đốc Phụng. Kết quả so sánh với bộ gene tham chiếu, bộ gene của giống lúa Đốc Phụng có khoảng 1.918.726 biến thể dạng thay đổi một nucleotide (Single Nucleotide Polymorphism) và chèn vào khoảng 81.435, mất đi khoảng 81.974. Trong khi đó ở giống Nép Mỡ, có khoảng 1.931.380 SNP và chèn vào khoảng 88.473, mất đi khoảng 83.190 vùng DNA. Đa số các biến thể xuất hiện ở các vùng không mang chức năng như trước sau và giữa các gene chiếm tỉ lệ trên 75%. Kết quả khảo sát biến thể xuất hiện trong vùng gene *OsTZF1* (LOC_Os05g10670.1), có chức năng điều hòa các nhóm gene liên quan đến các yếu tố stress sinh học và phi sinh học, cho thấy ở giống Đốc Phụng có 7 biến thể SNP và có chèn thêm 9 nucleotide mã hóa 3 amino acid arginine khi so với giống Nép Mỡ dựa trên bộ gene tham chiếu. Thông tin này giúp cho các nhà chọn giống sử dụng nó như chỉ thị phân tử, chọn tạo giống chống chịu mặn trong tương lai.

1. GIỚI THIỆU

Lúa (*Oryza sativa* L.) là cây trồng quan trọng trên thế giới, được trồng trên nhiều quốc gia, cung cấp lương thực chính cho 50% dân số trên thế giới (Khush, 1997; Islam et al., 2018) và có tính thương mại. Tuy nhiên, trước thực trạng của biến đổi khí hậu, hạn hán, xâm nhập mặn đã và đang là vấn đề cấp bách ảnh hưởng đến sản lượng lúa gạo, đặc biệt tác động nặng nề đến nền nông nghiệp Việt Nam, Đồng bằng sông Cửu Long (Nguyễn Thanh Tường và ctv., 2011), các tỉnh ven biển chịu sự xâm nhập mặn và ngập ngày càng gia tăng làm thay đổi đất canh tác lúa sang các vùng nuôi trồng thủy sản nước lợ (Phạm Thanh Vũ và ctv., 2016). Nhằm đảm bảo an ninh lương thực, việc canh tác giống lúa trên nền đất bị nhiễm mặn và đảm bảo sản lượng là cần thiết. Do đó, việc chọn tạo và khai thác nguồn gene lúa có khả năng chống chịu trong điều kiện đất nhiễm mặn đã và đang được các nhà khoa học quan tâm.

Khai thác và sử dụng nguồn gene lúa địa phương, ở những vùng bị xâm nhiễm mặn là một trong những chiến lược mà các nhà chọn tạo giống ưu tiên (Menguer et al., 2017; Rahman et al., 2016), tuy nhiên chiến lược này đôi khi không thành công vì nguồn gene lúa có khả năng chống chịu mặn không mang lại năng suất, hoặc phụ thuộc rất nhiều vào thời vụ và thường thì chất lượng gạo không cao (Rahman et al., 2016). Vì vậy, có rất nhiều báo cáo đã khai thác thông tin của bộ gene từ các giống lúa địa phương trong công tác chọn tạo giống lúa chịu mặn (Lakra et al., 2019; Subudhi et al., 2020; Yuan et al., 2020). Ngày nay với tiến bộ của khoa học công nghệ, kỹ thuật giải trình tự thế hệ mới (Next Generation Sequencing-NGS) đã phát triển mạnh mẽ giúp cho việc khai thác thông tin di truyền một cách nhanh chóng, hiệu quả và hỗ trợ cho công tác chọn tạo giống dựa trên nền tảng dấu chỉ thị phân tử (Barba et al., 2014). Bằng việc sử dụng kỹ thuật NGS đã xây dựng bộ gene tham chiếu trên cây lúa một cách hoàn chỉnh và chính xác hơn (Kawahara et al., 2013), từ cơ sở đó đã giúp cho cung cấp thông tin về các đa hình khi sử dụng bộ gene tham chiếu để so sánh các bộ gene cần nghiên cứu (Huang et al., 2013). Thực vậy, việc phát hiện ra các biến thể trên toàn bộ bộ gene tạo cơ hội để làm sáng tỏ cơ sở phân tử của sự khác biệt kiểu hình, đặc biệt là các biến thể giúp cho cây trồng có khả năng chống chịu lại với các điều kiện bất lợi của môi trường. Trên cơ sở đó, nghiên cứu này đã ứng dụng công nghệ NGS để giải mã bộ gene của 2 giống lúa địa phương vùng Đồng bằng Sông Cửu Long (ĐBSCL) 1 giống có khả năng chịu mặn và 1 giống mặn cảm với mặn nhằm tìm ra các biến thể ở các gene mang tính chống chịu mặn

đã được xác định từ các nghiên cứu trước (Jain et al., 2014; Chen et al., 2017), từ đó có thể sử dụng chúng như các chỉ thị phân tử dùng cho chọn giống chống chịu mặn trong tương lai.

2. PHƯƠNG TIỆN VÀ PHƯƠNG PHÁP

2.1. Phương tiện

Hai giống lúa địa phương có kiểu hình tương phản được dùng trong thí nghiệm này là Đốc Phụng (chống chịu mặn) và Nếp Mỡ (mặn cảm mặn) được lấy từ ngân hàng giống của Khoa Nông nghiệp, Trường Đại học Cần Thơ. Hai giống lúa Đốc Phụng và Nếp Mỡ được lựa chọn cho nghiên cứu dựa trên kết quả từ nghiên cứu của Tam (2019) và Tin et al. (2021) về tính chống chịu cũng như mặn cảm với mặn.

2.2. Phương pháp

2.2.1. Phương pháp phân lập và giải trình tự DNA

DNA của 2 giống lúa dùng cho nghiên cứu được ly trích theo quy trình Doyle (1990), sau khi ly trích thành công, DNA được kiểm tra chất lượng bằng Nanodrop (ThermoFisher, USA) và Bioanalyzer (Agilent, USA). Sau đó mẫu được đưa giải trình tự bằng hệ thống Illumina HiSeq 2500™ sử dụng kỹ thuật giải trình tự 100-bp từ hai đầu theo quy trình của chuẩn của Illumina (Illumina, San Diego, CA).

2.2.2. Phân tích bộ gene

Hầu hết các bước phân tích được thực hiện trên nền Linux Ubuntu 18.04.3 LTS (<https://ubuntu.com/download/desktop>). Sử dụng chương trình fastp V0.20.0 như một bộ tiền xử lý file FASTQ để kiểm tra chất lượng và các tính năng lọc dữ liệu, chương trình này chạy nhanh hơn 2-5 lần các so với các chương trình xử lý khác như Trimmomatic hay Cutadapt (Chen et al., 2018). Những trình tự được đọc sau khi được kiểm tra chất lượng được gắn vào bộ gene tham chiếu Os-Nipponbare-Reference-IRGSP-1.0 (Kawahara et al., 2013) có trên Ensembl Plants website (Bolser et al., 2016) bằng cách sử dụng phần mềm HISAT2 V2.1.0 (Kim et al., 2015; Keel & Snelling, 2018), và loại bỏ hoàn toàn các low-mapping quality (MAPQ<30) bằng công cụ SAMtools toolkit V1.9 (Li et al., 2009). Để nhận dạng các biến thể, các lỗi trùng lặp (duplicates) được loại bỏ khỏi file đã được điều chỉnh bằng công cụ Picard V2.18.7 (<http://broadinstitute.github.io/picard/>).

2.2.3. Xác định biến thể (variant calling)

Sau khi loại bỏ các lỗi do trùng lặp gây ra và sắp xếp lại các vùng xung quanh InDels, sau đó sử dụng

chương trình SAMtools toolkit V1.9 (Li et al., 2009) và BCFtools V1.9 (Li, 2011) cho việc phát hiện biến thể SNPs, InDels. Tiếp theo, loại bỏ các biến thể được dò tìm có chất lượng thấp (%QUAL<20), chỉ giữ lại các biến thể có chất lượng cho các phân tích sâu hơn bằng công cụ BCFtools V1.9 (Li, 2011). Theo Li (2014), việc loại bỏ các biến thể bị trùng lặp (overlapping) cùng với sự xuất hiện của các vùng low-mapping sẽ mang lại hiệu quả cao trong việc loại bỏ các dị hợp tử giả (heterozygotes). Ngoài ra, đối với các vị trí có độ nhiễu cao (low-complexity) sẽ được đánh dấu là DUST (Morgulis et al., 2006) chúng được loại bỏ bằng công cụ minimap toolkit V0.2 (<https://github.com/lh3/minimap>) (Li, 2016; Chu et al., 2017).

2.2.4. Phân tích số liệu

Sử dụng chương trình IGV (integrative genome viewer) truy cập tại trang web Integrative Genome Viewer (<https://www.igv.org>) để kiểm tra và so sánh những biến thể đột biến chèn hoặc mất đoạn (InDels) giữa các mẫu so với bộ gene tham chiếu

Bảng 1. Số liệu của 3 giống lúa được giải trình tự

Giống lúa	Số liệu read thô (bp)	Số liệu read (MAPQ30)	Số liệu read được ghép (bp)	Tỉ lệ read được ghé (%)	Chiều dài của read thô (bp)
Độc Phụng	37.948.246	25.288.216	25.586.354	67,42	150*2
Nếp Mỡ	41.837.368	27.352.488	26.602.508	65,87	150*2

MAPQ30: Mapping quality of 30

3.2. Nhận dạng SNPs và InDels

Qua phân tích kết biến thể di truyền gồm đột biến điểm thay đổi nucleotide (SNPs) và đột biến mất hay chèn nucleotide (InDels) giữa 2 giống lúa dựa trên bộ gene tham chiếu Nipponbare thu được tổng số SNP khoảng 1.918.726 và chèn vào khoảng 81.435, mất đi khoảng 81.974 ở giống Độc Phụng. Trong khi đó ở giống Nếp Mỡ có khoảng 1.931.380 SNP và chèn vào khoảng 88.473, mất đi khoảng 83.190 (Bảng 2). Mật độ xuất hiện các loại biến thể trung

Bảng 2. Số liệu SNPs và InDels của 2 giống lúa

Giống lúa	SNP			InDel		
	Tổng số	Mật độ SNPs/100 kb	Chèn	Xóa	Tổng số	Mật độ InDels/100 kb
Độc Phụng	1.918.726	515	87.435	81.974	169.409	45
Nếp Mỡ	1.931.380	518	88.473	83.190	171.663	46

Sự phân bố của các đa hình DNA xuất hiện trong suốt 12 nhiễm sắc thể (NST) của 2 giống lúa (Hình 1). Phần lớn số lượng SNPs và InDels của mỗi NST được tìm thấy tỉ lệ thuận với chiều dài của NST đó trên cả Độc Phụng và Nếp Mỡ. Hầu hết SNPs được

Nipponbare phiên bản 7.0 (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) và so với các giống với nhau ở các vùng Quantitative Trait Locus (QTL) liên quan đến khả năng chống chịu mặn đã được Naveed et al. (2018) tìm thấy, từ đó để cho thấy sự khác biệt trình tự protein giữa giống chống chịu và giống mặn cảm.

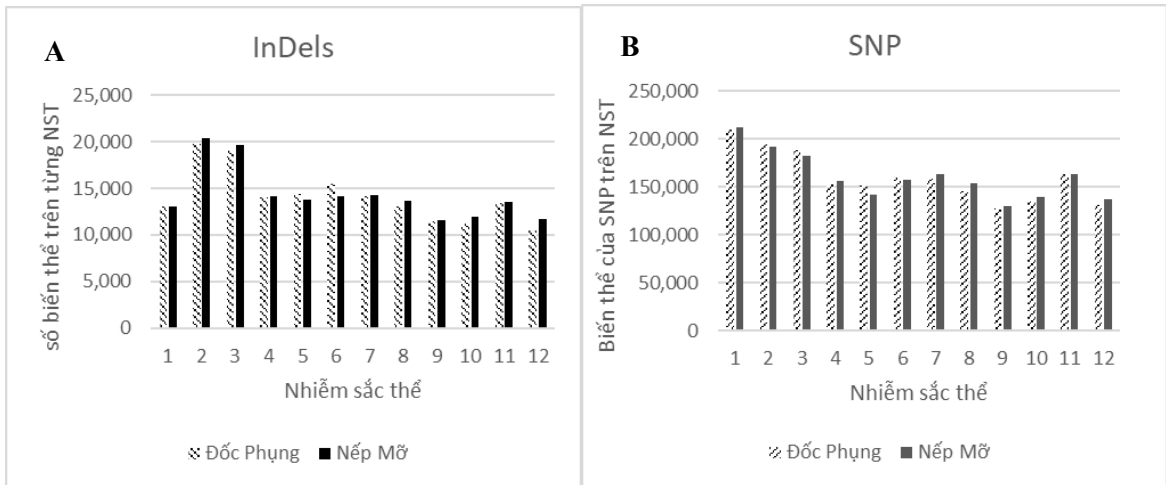
3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Kết quả đọc và ghép vào bộ gene tham chiếu

Kết quả giải trình tự bộ gene của 2 giống lúa thí nghiệm bằng kỹ thuật giải trình tự thế hệ mới thông qua hệ thống Illumina HiSeq 2500™ thu được số lượng đọc sau khi loại bỏ các lần đọc có chất lượng thấp hơn 30% tương ứng là 25.288.216 bp cho Độc Phụng và 27.352.488 bp cho Nếp Mỡ (Bảng 1). Tỉ lệ đọc được ghép hoàn chỉnh vào bộ gene Nipponbare đạt khoảng từ 65% - 67% của 2 giống lúa.

binh trong 100 kb cho SNP khoảng 515 và 518, tương ứng cho Độc Phụng và Nếp Mỡ. Mật độ xuất hiện cho Indel ở 2 giống này trong vùng 100 kb tương ứng 45 (Độc Phụng) và 46 (Nếp Mỡ). Kết quả nghiên cứu này cũng tương ứng như kết quả tìm thấy khi so sánh bộ gene của hai giống lúa hoang của Trung Quốc với bộ gene tham chiếu Nipponbare và 93-11 (Liu et al., 2017), và cũng được tìm thấy khi nghiên cứu bộ gene của 1.143 dòng lúa lai có nguồn gốc từ các bố mẹ ưu tú ở Trung Quốc (Lv et al., 2020).

phát hiện nhiều ở NST 1, 2, 3 và InDels phổ biến nhiều nhất ở NST 2 và 3. Tuy nhiên, các đa hình DNA đều xuất hiện ít ở NST 9, 10 và 12. Bên cạnh đó, số lượng biến thể SNPs và InDels của Nếp Mỡ chiếm cao hơn so với Độc Phụng.



Hình 1. Đa hình DNA trên 12 NST của ở các giống Đốc Phụng và Nếp Mỡ

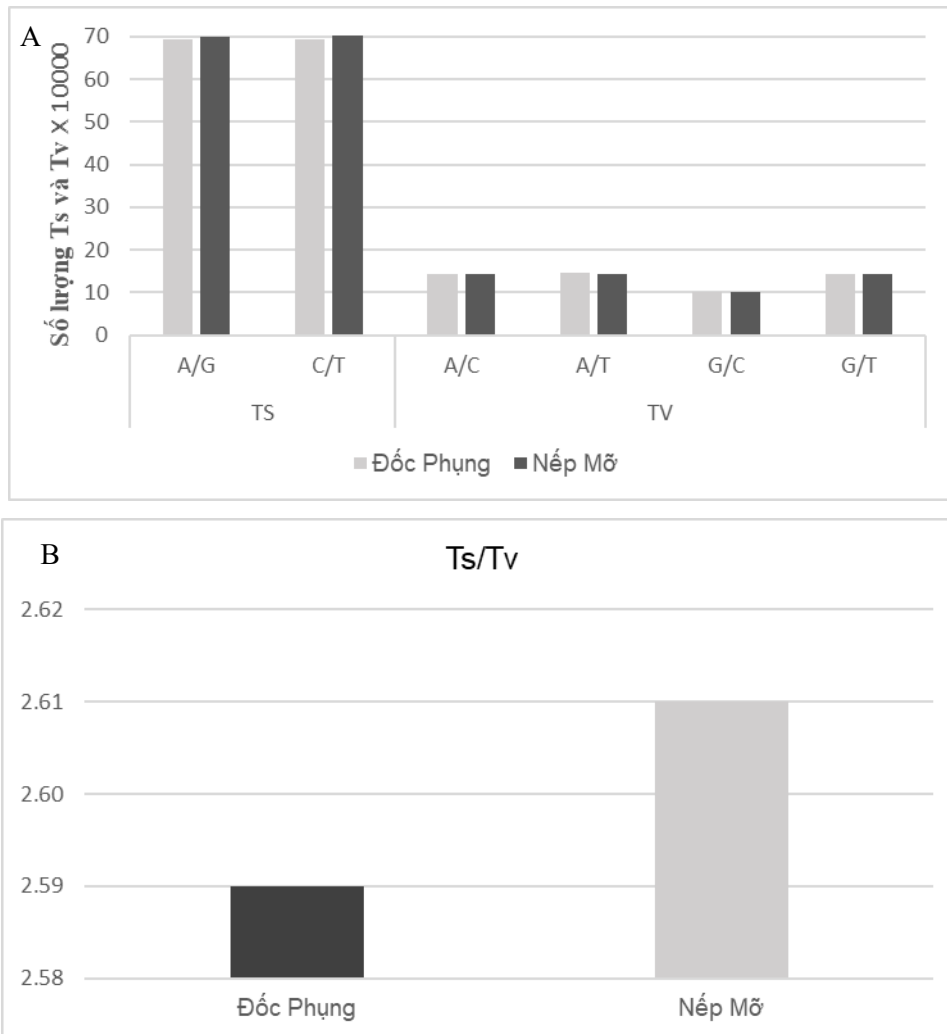
A. InDel, B. SNP

3.2.1. Phân tích SNPs và InDels

Tổng số đột biến đồng vị (transition -Ts) và đột biến chuyển vị (transversion- Tv) của 2 giống được trình bày trong Hình 2. Tần số của các Ts (A/G và C/T) ở Đốc Phụng, Nếp Mỡ tương tự nhau và đều cao hơn so với tần số Tv (A/C, A/T, C/G và G/T). Tần số A/G và C/T tương tự nhau trong mỗi giống (Hình 2A). Tuy nhiên, tần số của Tv không đồng đều, tần số của G/C thấp hơn so với ba loại Tv khác. Tỷ lệ Ts/Tv của Nếp Mỡ (2.617) cao hơn so với Đốc Phụng (2.593) (Hình 2B), tỷ lệ Ts/Tv được báo cáo trên lúa và ngô (Batley et al., 2003; Morton, 1995), tần số đột biến của Ts cao hơn so với Tv (thay đổi cùng dạng vòng sẽ mang nhiều lợi thế hơn) và khả năng thay đổi cấu trúc và chức năng protein của Ts sẽ thấp hơn so với Tv (Wakeley, 1996; Subbaiyan et al., 2012). Như vậy, với tỷ lệ Ts/Tv càng thấp thì khả năng đột biến có ảnh hưởng lên protein càng lớn, với kết quả này thì giống Đốc Phụng có khả năng biến đổi do đột biến cao hơn Nếp Mỡ.

Một trong những đặc điểm quan trọng của đột biến là vị trí của đa hình DNA vì được biết có ảnh hưởng đến sự biểu hiện và chức năng gene (Subbaiyan et al., 2012; Subudhi et al., 2020). Các đa hình DNA xuất hiện trong vùng mã hóa là rất

quan trọng vì chúng có thể làm thay đổi chức năng của protein, các biến thể xuất hiện trong vùng enhancer có thể làm giảm hay làm tăng sự biểu hiện gene. Do đó, việc phát hiện ra các đa hình liên quan đến sự thay đổi chức năng của gene là rất quan trọng để nghiên cứu về sự khác biệt của các kiểu hình (Jain, 2012; Jain et al., 2014). Nhìn chung, sự phân bố của SNPs và InDels ở các vùng gene khác nhau có tỷ lệ tương đương cho các so sánh của 2 giống. SNPs và InDels xuất hiện nhiều ở các vùng không mã hóa (bao gồm các vùng giữa (intergenic), trên (upstream), dưới (downstream) các gene) so với các vùng mã hóa (CDS) (Hình 3). Tần suất của các đột biến di truyền xuất hiện nhiều ở các vùng không mã hóa có thể do chịu áp lực ít hơn từ chọn lọc tự nhiên (Barreiro et al., 2008). Cụ thể, các đa hình DNA xuất hiện nhiều nhất trong vùng upstream và downstream chiếm tỷ lệ gần bằng và hơn 30%, tiếp đến là vùng intergenic đạt hơn 21% và 16% lần lượt cho các biến thể SNPs và InDels, vùng intron đạt khoảng 3,1-4,4%, trong khi đó vùng mã hóa chỉ chiếm tỷ lệ khoảng 1,6-2,9%, cuối cùng gồm các vùng exon, 5'UTR và 3'UTR có tỷ lệ thấp nhất chỉ khoảng 0,4-1,1% (Hình 3).

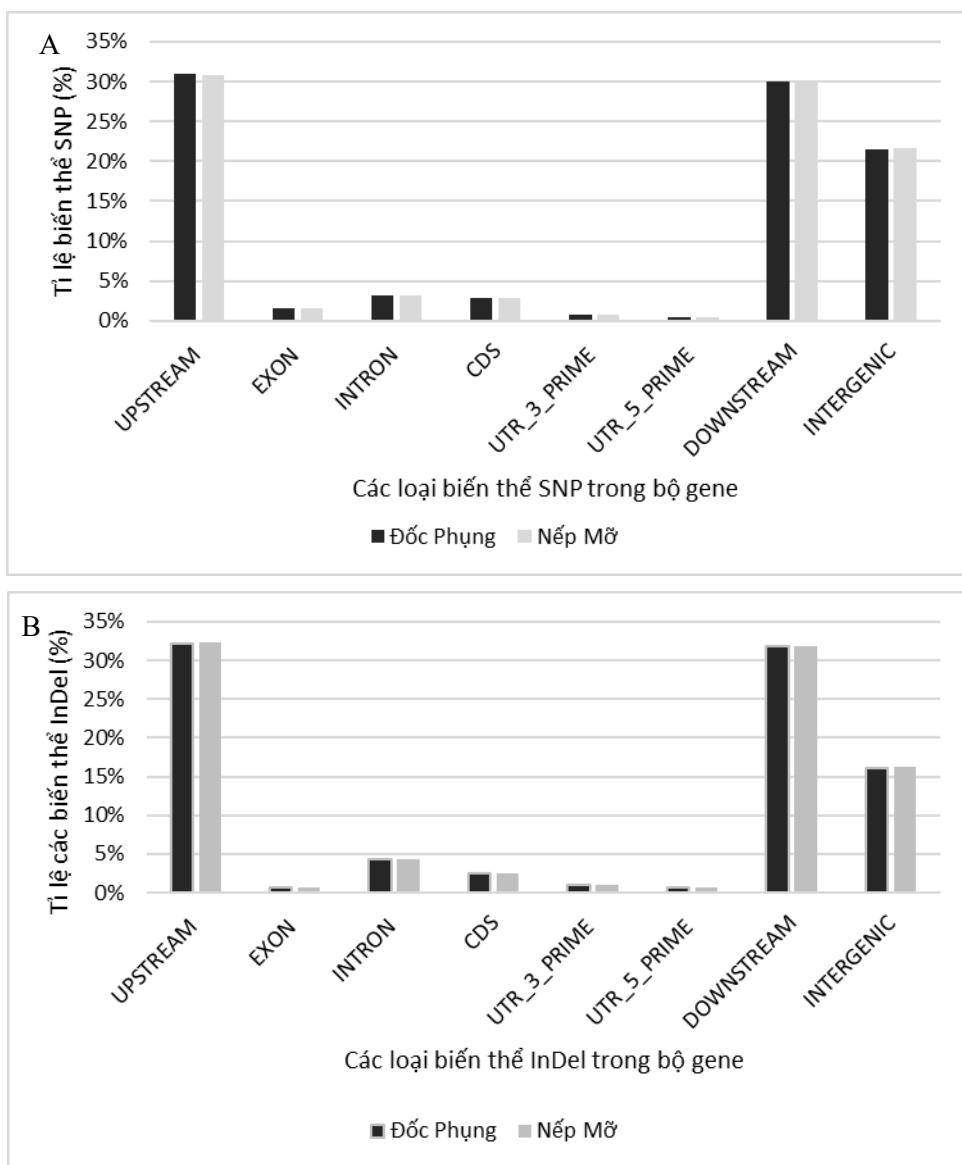


Hình 2. Phân loại biến thể ở hai giống lúa Đốc Phụng và Nếp Mỡ

A. số lượng loại biến thể Ts và Tv; B. Tỷ lệ biến thể Ts/Tv

Các biến thể xảy ra bên ngoài vùng mã hóa cũng có khả năng tác động đến sự biểu hiện của gene như đột biến ở vùng promoter, các vị trí ghép nối các exon trong quá trình tổng hợp RNA thông tin trưởng thành (splice-site regions), hay các vị trí liên kết của các ribosome. Bên cạnh đó, các đột biến liên quan đến trình tự DNA như đột biến sai nghĩa (missense), đột biến vô nghĩa (nonsense) hay đột biến im lặng (silent) cũng tác động đến quá trình tổng hợp protein, trong đó dạng đột biến sai nghĩa có tính đóng góp hơn so với 2 dạng đột biến còn lại, nó làm

thay đổi acid amin từ đó có thể ảnh hưởng hoặc không đến biểu hiện chức năng của protein, trong khi đột biến im lặng không làm thay đổi amino acid, tuy nhiên trong một số trường hợp chúng vẫn ảnh hưởng đến sự biểu hiện của kiểu hình bằng cách tăng hay làm chậm quá trình tổng hợp protein. Tỷ lệ đột biến của 2 giống lúa được thể hiện trong Bảng 3, hầu hết tỷ lệ của đột biến sai nghĩa chiếm hơn 50% trong tổng số các dạng đột biến có mức ảnh hưởng đến chức năng, tuy nhiên tỷ lệ đột biến này ở giống Nếp Mỡ thấp hơn so với Đốc Phụng.



Hình 3. Tỷ lệ SNP (A) và InDels (B) xuất hiện ở các vị trí trong bộ gene

Bảng 3. Tỷ lệ các loại đột biến xuất hiện trên 2 giống lúa thí nghiệm

Tên giống	Tỷ lệ đột biến (%)		
	Sai nghĩa (Missense)	Vô nghĩa (Nonsense)	Im lặng (Silent)
Đốc Phụng	52,86	0,75	46,39
Nếp Mỡ	52,73	0,76	46,51

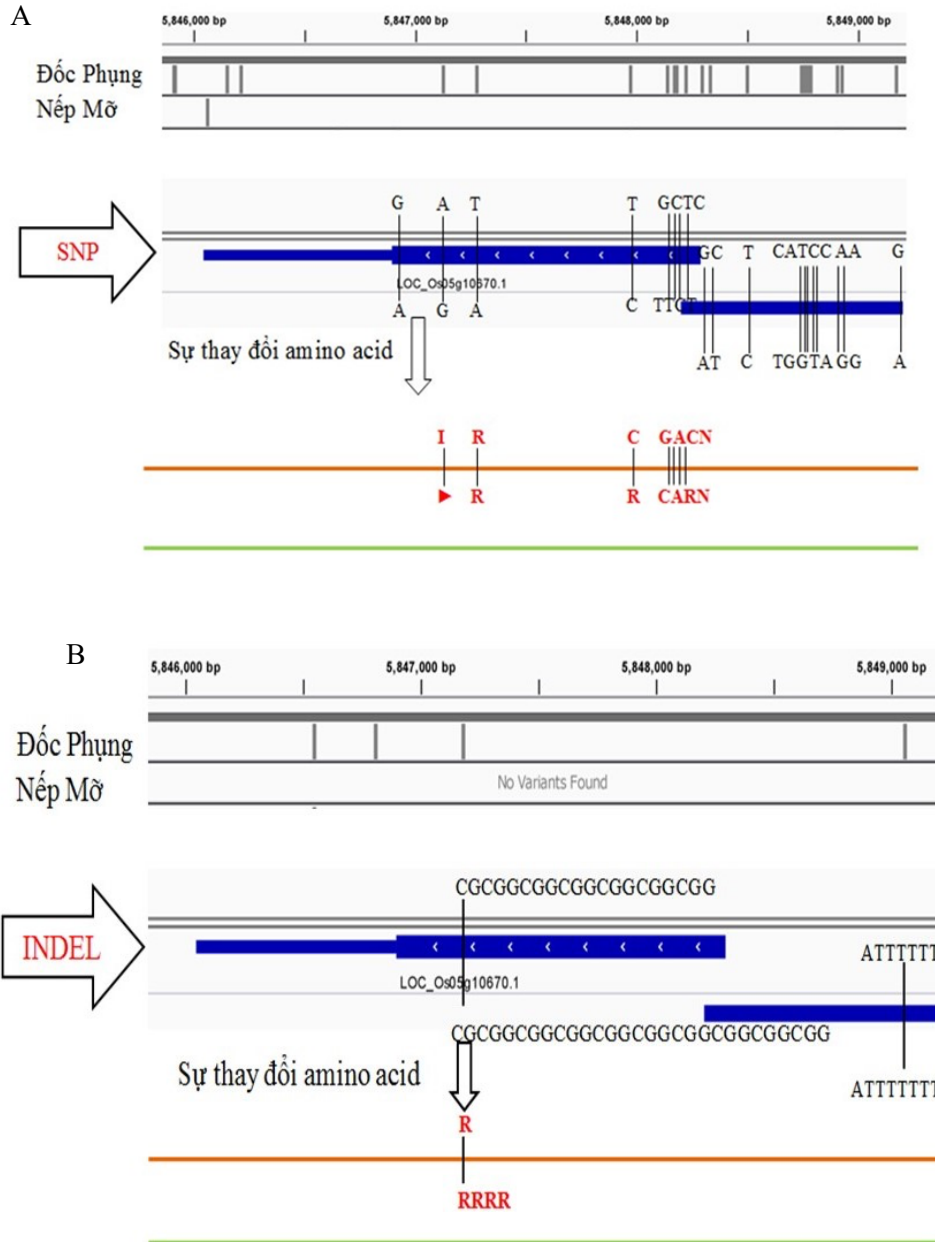
3.2.2. Phân tích biến thể có trong các gene chịu mặn trên lúa

Phần lớn các biến thể xuất hiện nhiều ở vùng intergenic, promoter và ít hơn ở vùng mã hóa DNA (Fuentes et al., 2019). Sự hiện diện của các biến thể ở vùng mã hóa exon được chú trọng hơn so với các vùng khác, mặc dù các biến thể trong vùng này

không đủ để thiết lập bản đồ cho các đặc tính nông học phức tạp (Steri et al., 2018). Biến thể SNP/InDel hiện diện trên vùng promoter hay sự liên kết giữa các biến thể vùng này và vùng mã hóa đóng vai trò quan trọng ảnh hưởng đến các đặc điểm nông học trên cây trồng, tuy nhiên các hiểu biết về sự ảnh hưởng này vẫn chưa được hiểu rõ và chưa có nhiều báo cáo trên cây lúa (Wang et al., 2018; Wang et al.,

2019). Ngoài ra, các biến thể di truyền tại vùng unstranlated region (UTR) có thể thay đổi các yếu tố điều hòa từ đó tác động đến quá trình phiên mã của mRNA (Steri et al., 2018). Đáng chú ý hơn là có đến 3,7% biến thể xuất hiện ở vùng UTR theo các

nghiên cứu liên quan đến Genome-wide association studies (GWAS) (Hindorff et al., 2009). Nhìn chung, mức độ phiên mã của gene có thể bị ảnh hưởng bởi SNP và InDel (Kreimer et al., 2012).



Hình 4. Biến thể SNP (A) và InDel mất đi 3 amino acid Arginine (R) (B) ở giống lúa Độc Phụng so với Nếp Mỡ

Trong nghiên cứu này, chương trình IGV được sử dụng để phát hiện vị trí của SNPs và InDels ở các gene có liên quan đến khả năng chống chịu mặn trên

12 NST đã được báo cáo trước đây (Naveed et al., 2018), từ đó so sánh sự khác nhau của biến thể trên cùng một gene của 2 giống lúa. Đặc biệt trong đó

biến thể được phát hiện ở gene *OstZF1* (LOC_Os05g10670.1) mang chức năng điều hòa hậu phiên mã của nhóm gene phản ứng với stress sinh học và phi sinh học (stress với hạn, mặn cao) (Jan et al., 2013; Seong et al., 2020). Từ những nghiên cứu trên, ảnh hưởng của SNPs và InDel có ở gene này đã ảnh hưởng đến khả năng biểu hiện chức năng của gene rất quan trọng dưới tác động của mặn.

Khảo sát trên NST số 5 tại vùng gene *OstZF1* (LOC_Os05g10670.1) của giống lúa Đốc Phụng cho thấy biến thể SNPs (Hình 4A) và InDels (Hình 4B) thể hiện sự khác biệt rõ ràng với giống Nếp Mỡ. Hình 4A cho thấy vùng gene này có sự hiện diện của đa hình ở Đốc Phụng là 16 vị trí (vùng exon: 7 vị trí) so với Nếp Mỡ chỉ có 1 vị trí khi so sánh với bộ gene tham chiếu. Ngoài ra, Đốc Phụng có biến thể InDels trong vùng mã hóa exon tại vị trí 5.847.161-5.847.180 bp đã làm thay đổi chuỗi polypeptide bằng cách chèn thêm 3 amino acid arginine, trong khi đó không có ở giống Nếp Mỡ (Hình 4B). Như vậy, biến thể ở gene này có thể liên quan đến việc điều hòa nhóm gene giúp cho giống Đốc Phụng chống chịu tốt với stress mặn, như những nghiên cứu trước (Jan et al., 2013; Naveed et al., 2018; Seong et al., 2020). Kết quả nghiên cứu này cho thấy có thể sử dụng đột biến mất 3 amino acid arginine của gene *OstZF1* được xem như là dấu chỉ thị phân tử là gene chức năng liên quan đến khả năng chống chịu mặn ở giống lúa Đốc Phụng nhằm phục vụ cho công tác chọn tạo giống lúa chống chịu mặn trong tương lai.

4. KẾT LUẬN

Tóm lại, nghiên cứu này cho thấy phương pháp tiếp cận gene chức năng dựa trên cơ sở giải trình tự bộ gene là một trong những kỹ thuật hiệu quả nhất để xác định các gene chứa các biến thể SNP và InDel liên quan đến các đặc điểm mong muốn. Việc phân tích biến thể xuất hiện ở giống lúa chịu mặn Đốc Phụng ở gene mục tiêu *OstZF1* là gene ứng viên cho điều hòa hoạt động các gene liên quan đến đặc tính chống chịu stress mặn là hợp lý. Tuy nhiên, cần có thêm bằng chứng để xác nhận chức năng của các gene này; chỉnh sửa bộ gene bằng một phương pháp hợp lý sẽ là cách tốt nhất để thực hiện điều này. Do đó, đề nghị tiếp tục dùng phương pháp chỉnh sửa bộ gene và biểu hiện gene để có kết luận chính xác về chức năng của gene này trên cây lúa.

LỜI CẢM ƠN

Nghiên cứu được tài trợ bởi dự án Nâng cấp Trường Đại học Cần Thơ VN14-P6 (vốn vay ODA từ Chính phủ Nhật Bản).

TÀI LIỆU THAM KHẢO

- Barba, M., Czosnek, H., & Hadidi, A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses*, 6(1), 106-136.
- Barreiro, L. B., Laval, G., Quach, H., Patin, E., & Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nature Genetics*, 40, 340-345.
- Batley, J., Barker, G., O'Sullivan, H., Edwards, K. J., & Edwards, D. (2003). Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant physiology*, 132, 84-91.
- Bolser, D., Staines, D. M., Pritchard, E., & Kersey, P. (2016). Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. In: Edwards, D. (Ed.), *Plant Bioinformatics: Methods and Protocols*. Springer New York, New York, NY, pp. 115-140.
- Chen, R., Cheng, Y., Han, S., Van Handel, B., Dong, L., Li, X., & Xie, X. (2017). Whole genome sequencing and comparative transcriptome analysis of a novel seawater adapted, salt-resistant rice cultivar – sea rice 86. *BMC Genomics*, 18, 655.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34, i884-i890.
- Chu, J., Mohamadi, H., Warren, R. L., Yang, C., & Birol, I. (2017). Innovations and challenges in detecting long read overlaps: an evaluation of the state-of-the-art. *Bioinformatics (Oxford, England)*, 33, 1261-1270.
- Doyle, J. J., and Doyle, J. J. (1990). Isolation of plant DNA from fresh tissue. *Focus*, 12, 13-15.
- Fuentes, R. R., Chebotarov, D., Duitama, J., Smith, S., De la Hoz, J. F., Mohiyuddin, M., Wing, R. A., McNally, K. L., Tatarinova, T., Grigoriev, A., Mauleon, R., & Alexandrov, N. (2019). Structural variants in 3000 rice genomes. *Genome Research*, 29, 870-880.
- Hindorf, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106, 9362-9367.
- Huang, X., Lu, T., & Han, B. (2013). Resequencing rice genomes: an emerging new era of rice genomics. *Trends in genetics: TIG*, 29, 225-232.
- Islam, M. Z., Khalequzzaman, M., Prince, M., Siddique, M. A., Rashid, E., Ahmed, M., Pittendrigh, B. R., & Ali, M. P. (2018). Diversity

- and population structure of red rice germplasm in Bangladesh. *PLOS ONE*, 13, e0196096.
- Jain, M. (2012). Next-generation sequencing technologies for gene expression profiling in plants. *Brief Funct Genomics*, 11, 63-70.
- Jain, M., Moharana, K. C., Shankar, R., Kumari, R., & Garg, R. (2014). Genome-wide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. *Plant Biotechnology Journal*, 12, 253-264.
- Jan, A., Maruyama, K., Todaka, D., Kidokoro, S., Abo, M., Yoshimura, E., Shinozaki, K., Nakashima, K., & Yamaguchi-Shinozaki, K. (2013). OsTZF1, a CCH-Tandem Zinc Finger Protein, Confers Delayed Senescence and Stress Tolerance in Rice by Regulating Stress-Related Genes. *Plant Physiology*, 161, 1202-1216.
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., ... & Matsumoto, T. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, 6(1), 1-10.
- Keel, B. N., and Snelling, W. M. (2018). Comparison of Burrows-Wheeler Transform-Based Mapping Algorithms Used in High-Throughput Whole-Genome Sequencing: Application to Illumina Data for Livestock Genomes. *Frontiers in genetics*, 9, 35-35.
- Khush, G. S. (1997). Origin, dispersal, cultivation and variation of rice. Springer, Dordrecht.
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12, 357-360.
- Kreimer, A., Litvin, O., Hao, K., Molony, C., Pe'er, D., & Pe'er, I. (2012). Inference of modules associated to eQTLs. *Nucleic acids research*, 40(13), e98-e98.
- Lakra, N., Kaur, C., Singla-Pareek, S. L., & Pareek, A. (2019). Mapping the 'early salinity response' triggered proteome adaptation in contrasting rice genotypes using iTRAQ approach. *Rice*, 12(1), 1-22.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987-2993.
- Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, 30(20), 2843-2851.
- Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, 32(14), 2103-2110.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- Liu, W., Ghouri, F., Yu, H., Li, X., Yu, S., Shahid, M. Q., & Liu, X. (2017). Genome wide re-sequencing of newly developed Rice Lines from common wild rice (*Oryza rufipogon* Griff.) for the identification of NBS-LRR genes. *PLOS ONE*, 12, e0180662.
- Lv, Q., Li, W., Sun, Z., Ouyang, N., Jing, X., He, Q., ... & Yuan, D. (2020). Resequencing of 1,143 indica rice accessions reveals important genetic variations and different heterosis patterns. *Nature communications*, 11(1), 1-10.
- Menguer, P. K., Sperotto, R. A., & Ricachenevsky, F. K. (2017). A walk on the wild side: *Oryza* species as source for rice abiotic stress tolerance. *Genetics and molecular biology*, 40, 238-252.
- Morgulis, A., Gertz, E. M., Schäffer, A. A., & Agarwala, R. (2006). A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *Journal of Computational Biology*, 13(5), 1028-1040.
- Morton, B. R. (1995). Neighboring base composition and transversion/transition bias in a comparison of rice and maize chloroplast noncoding regions. *Proceedings of the National Academy of Sciences*, 92(21), 9717-9721.
- Naveed, S. A., Zhang, F., Zhang, J., Zheng, T. Q., Meng, L. J., Pang, Y. L., ... & Li, Z. K. (2018). Identification of QTN and candidate genes for salinity tolerance at the germination and seedling stages in rice by genome-wide association analyses. *Scientific reports*, 8(1), 1-11.
- Nguyễn Thanh Tường, Nguyễn Tâm Đạo, Võ Công Thành & Nguyễn Bảo Vệ. (2011). Tuyển chọn giống lúa cao sản chịu mặn cho vùng lúa – tôm tại tỉnh Bạc Liêu. *Tạp chí Nông nghiệp và Phát triển Nông thôn*, 12, 30-36.
- Phạm Thanh Vũ, Võ Quang Minh, Vương Tuấn Huy và Phan Chí Nguyên. (2016). Tác động của mặn và ngập theo kịch bản biến đổi khí hậu đến tiềm năng thích nghi đất đai vùng ven biển Đồng bằng sông Cửu Long. *Tạp chí Khoa học Trường Đại học Cần Thơ, Số chuyên đề: Nông nghiệp*, 71-83.
- Rahman, M. A., Thomson, M. J., Shah-E-Alam, M., de Ocampo, M., Egdane, J., & Ismail, A. M. (2016). Exploring novel genetic sources of salinity tolerance in rice through molecular and physiological characterization. *Annals of botany*, 117(6), 1083-1097.

- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, 29(1), 24-26.
- Seong, S. Y., Shim, J. S., Bang, S. W., & Kim, J. K. (2020). Overexpression of OsC3H10, a CCCH-zinc finger, improves drought tolerance in rice by regulating stress-related genes. *Plants*, 9(10), 1298.
- Steri, M., Idda, M. L., Whalen, M. B., & Orrù, V. (2018). Genetic variants in mRNA untranslated regions. *Wiley Interdisciplinary Reviews: RNA*, 9(4), e1474.
- Subbaiyan, G. K., Waters, D. L. E., Katiyar, S. K., Sadananda, A. R., Vaddadi, S., & Henry, R. J. (2012). Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnology Journal*, 10, 623-634.
- Subudhi, P. K., Shankar, R., & Jain, M. (2020). Whole genome sequence analysis of rice genotypes with contrasting response to salinity stress. *Scientific Reports*, 10, 21259.
- Tam, N. T. (2019). The genome constitution of rice resources in the Mekong Delta and their association with salinity stress (Doctoral dissertation). Hokkaido University.
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, 14(2), 178-192.
- Tim, H. Q., Loi, N. H., Labarosa, S. J. E., McNally, K. L., McCouch, S., & Kilian, B. (2021). Phenotypic response of farmer-selected CWR-derived rice lines to salt stress in the Mekong Delta. *Crop Science*, 61(1), 201-218.
- Wakeley, J. (1996). The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends in ecology & evolution*, 11(4), 158-162.
- Wang, J., Zhu, J., Zhang, Y., Fan, F., Li, W., Wang, F., ... & Yang, J. (2018). Comparative transcriptome analysis reveals molecular response to salinity stress of salt-tolerant and sensitive genotypes of indica rice at seedling stage. *Scientific reports*, 8(1), 1-13.
- Wang, P., Xiong, Y., Gong, R., Yang, Y., Fan, K., & Yu, S. (2019). A key variant in the cis-regulatory element of flowering gene Ghd8 associated with cold tolerance in rice. *Scientific reports*, 9(1), 1-14.
- Yuan, J., Wang, X., Zhao, Y., Khan, N. U., Zhao, Z., Zhang, Y., ... & Li, Z. (2020). Genetic basis and identification of candidate genes for salt tolerance in rice by GWAS. *Scientific reports*, 10(1), 1-9.