The genome constitution of rice resources in the Mekong Delta and their association with salinity stress

（メコンデルタにおけるイネ栽培系統集団のゲノム構成と塩ストレスとの関連性）

北海道大学　大学院農学院
生物資源科学専攻　博士後期課程

Nguyen Thanh Tam

The genome constitution of rice resources in the Mekong Delta and their association with salinity stress

Graduate School of Agriculture – Hokkaido University
Division of Agrobiology

Nguyen Thanh Tam

ABBREVIATIONS

Chr	Chromosome
DAT	Days after treatment
GJ	Japonica
GJ-tmp	Primarily East Asian temperate
GJ-sbtrp	Southeast Asian subtropical
GJ-trp	Southeast Asian tropical
GLM	General linear model
GWAS	Genome-wide association study
IRRI	International Rice Research Institute
LD	Linkage disequilibrium
MABC	Marker-assisted backcrossing
MAF	Minor allele frequency
MAS	Marker-assisted selection
MDI	Mekong Delta Development Institute
PCA	Principal component analysis
PHE	Plant height elongation
QTL	Quantitative
ddRAD-seq	digest restriction-site associated DNA sequencing
RLE	Root length elongation
RW	Root dry weight
SES	Standard evaluation system
SNP	Single nucleotide polymorphism
SW	Shoot dry weight
TASSEL	Trait Analysis by Association Evolution and Linkage
UTR	Untranslated region

CONTENTS

Chapter 1. General Introduction

1.1. The role of rice production

Rice (*Oryza sativa* L.) farming is the principal agricultural activity in many countries around the world, particularly South and Southeast Asia countries. Rice is one of staple foods because it provides up to 50% of the dietary calories of Asian people (Calpe, 2006). Nearly half of the world's population and 90% of Asian people rely on rice every day (Hariadi et al., 2015). Rice is grown in the areas where there is access to water resource (Batayeva et al., 2018). Most of rice production occurs in Asia which accounts for 90% of world rice market. Recently, rice is also becoming an important cereal food for many people in Central America, Europe and Africa (Muthayya et al., 2014).

Global rice consumption was approximately 490 million metric tons in 2019. Three top rice consumer countries from Asia were China (about 143.8 million metric tons), India (97.35 million metric tons), and Indonesia (37.4 million metric tons). Although many countries produce rice for self-consumption, there are some countries that export to other countries to sustain food security in the world. The first rice exporter country worldwide is India (12.5 million metric tons), followed by Thailand (10.3 million metric tons) and Vietnam (7.0 million metric tons) in 2019 (https://www.statista.com/statistics/255947/top-rice-exporting-countries-worldwide-2011/-21-5-2019).

In Vietnam, rice produced in the Mekong Delta contributes to 54% of the total rice productivity and 90% of the total export volume of Vietnam (Thanh et al., 2013). This contributes to 19% of the world's rice market (Nations, 2015).

1.2. Rice cultivars diversity

Rice cultivars have a wide variation in the worldwide with a known two common species such as *Oryza sativa* and *Oryza glaberrima* (Evenson & Gollin, 1994, Chang, 2003). *O. sativa* consists of two subgroups, *japonica* and *indica* (Zhao et al., 2010).

More than 40,000 varieties of *O. sativa* are cultivated widely in Asia countries. *O. glaberrima* is grown in most Africa countries (Vaughan et al., 2008). The diversity of rice can be identified based on their phenotypic differences in traits such as grain length, color, thickness, stickiness, aroma, growing habit, and others, as well as their genetic differences (Chang, 2003). Their diversity can also be seen on their morphological characters and in their physiological mechanisms correlated to the adaptation or specialization of ecosystem (Chang, 2003).

Nine rice subpopulations were classified from 3,010 rice genome accessions that were collected from 89 countries. Comparing with geographic origins, four XI clusters were identified that were XI-1A from East Asia, XI-1B as modern varieties of diverse origins, XI-2 from South Asia and XI-3 from Southeast Asia; three GJ clusters were formed and these were GJ-tmp (primarily East Asian temperate), GJ-sbtrp (Southeast Asian subtropical) and GJ-trp (Southeast Asian tropical); and single groups for the mostly South Asian cA and cB accessions were identified. Moreover, accessions with admixture components within XI (indica) were named XI-adm' and GJ (japonica) were 'GJ-adm', and lastly was admixed group for accessions that fell between these major groups (Wang et al., 2018).

In other classification, IRRI (2014) classified rice accessions into 12 subpopulations from 3024 variety names: ind1a, ind1b, ind2, ind3, indx, japx, temp, trop, subtrop, aro, aus, admix. Although the name of accessions group is different between IRRI (2014) and Wang et al. (2018), the character of each group is totally the same. Ind (indica rice accessions) was replaced for XI, trop was for GJ-trp, subtrop was for GJ-sbtrop, aro for cB, aus for CA. These subgroup names are the ones used in the SNP-seek database.

Knowledge on genetic diversity is the foundation of the genetic improvement of crops (Chang, 2003). Therefore, understanding rice genetic diversity will help rice breeders benefit on breeding and releasing new rice varieties that can adapt to climate change, resistant to pests and diseases, and other unfavorable conditions, especially, high grain yield or good quality.

1.3. Rice salinity tolerant studies

Rice is very important for agriculture and economics worldwide, but rice production deal with many issues, one of them is salt stress in coastal areas (Reddy et al., 2017). During dry season, salt will intrude from the sea to inland. Most of rice accessions cannot grow normally where salt concentration is above 3 dSm^{-1} (Hoang et al., 2015), especially at seedling stage (Batayeva et al., 2018), and reproductive stage (Quan et al., 2018). Improving rice production is urgent requirement to meet the food demand due to population increase. Unfortunately, the rice growth land cannot increase, whereas rice land will decrease with the increase of population and be affected by climate change. Therefore, research on salinity tolerance of rice cultivars is very important for agriculture worldwide. Up to now, salinity tolerance studies have been done on three major approaches: (i) conventional breeding, (ii) marker assisted selection and (iii) genetic engineering (Hoang et al., 2015).

Conventional breeding is the method for improving or developing of rice cultivars by conservation on plant genome within the natural genetic boundaries of the species (Acquaah, 2015, Gilliham et al., 2017). The purpose of this approach is to improve yield, the quality of crop product, the agronomic suitability and the resistant to important conditions, but focused to improve yields in normal conditions rather than stress conditions (Gilliham et al., 2017). To achieve success in breeding, the breeder applies general steps that will result into objectives accomplishment; these steps include creating and/or assembly of variabilities, selection, evaluation and finally cultivar release (Acquaah, 2015).

Marker assisted selection (MAS) for salinity tolerance study was applied from many previous studies. MAS is the process of using DNA markers as indirect selection criteria for selecting agriculturally important traits in crop breeding, which is used to improve the effectiveness or efficiency of selection for the traits of interest in breeding programs (Ashraf et al., 2012). The basic procedure for using MAS in crop breeding for salt tolerance is as follows; (1) identification of parental lines contrasting in salt tolerance; (2) development of mapping population; (3) identification of polymorphic genetic markers between the parental lines; (4) development of genetic maps; (5)

screening population for salt tolerance; (6) QTL analysis and identification of markers associated with salt tolerance; (7) conducting MAS; (8) combine MAS with polymorphic. To make MAS successful, large populations size and high-throughput markers are needed (Ashraf et al., 2012).

In MAS, QTL analysis has proven to be useful for the identification of genes responsible for biotic and abiotic stress tolerance in crops, therefore, genes responsible for salinity tolerance were identified in rice (Mittler & Blumwald, 2010). There are many previous studies which applied QTL analysis to identify the genes or genomic regions for salinity tolerance in rice cultivars (Flowers et al., 2000, Koyama et al., 2001, Lin et al., 2004, Lee et al., 2007, Singh et al., 2007, Thomson et al., 2010, Islam et al., 2011). Applying marker-assisted backcrossing (MABC) has created new salt tolerance rice cultivars (Linh le et al., 2012). Genetic engineering has been successfully used to characterize and transfer genes responsible for biosynthesis of different metabolites, e.g., proline, trehalose and polyamines from different organisms to crop plants to achieve the targeted approach (Mitra, 2001). Nowadays, the scientists used genome-editing approach for their breeding strategies.

Up to now, 87 genes were known for salinity tolerance in 12 chromosomes of rice genome (Fig. 1.1). The highest number of salinity trait genes were in Chr01 (16 genes), followed by Chr02 (15 genes).

Currently, about 2,000 samples of local and improved rice varieties in the Mekong Delta have been collected and conserved at the Mekong Delta Research Development Institute of Can Tho University (Viet Nam) as the genetic material for salinity-tolerant rice breeding. Therefore, the evaluation and classification of rice varieties is one of the solutions to deal with projected climate change in the future.

1.4. Objectives

This study focused on 99 *indica* rice accessions (81 landraces) collected from 10 provinces in the coastal region of Mekong Delta in Vietnam to evaluate genomic

structure for local bioresources. I investigated genetic relationships between Mekong Delta and neighboring countries, and genomic contribution of the landraces to the improved varieties in the Mekong Delta rice accessions. This study also aimed to screen and evaluate a set of MDI rice accessions for salinity tolerance at seedling stage. Moreover, I tried to identify the relationship between phenotypes on four plant characteristics of salt tolerance traits (plant height, root length, shoot dry weight and root dry weight) and SNPs from RAD-seq and whole genome sequence.

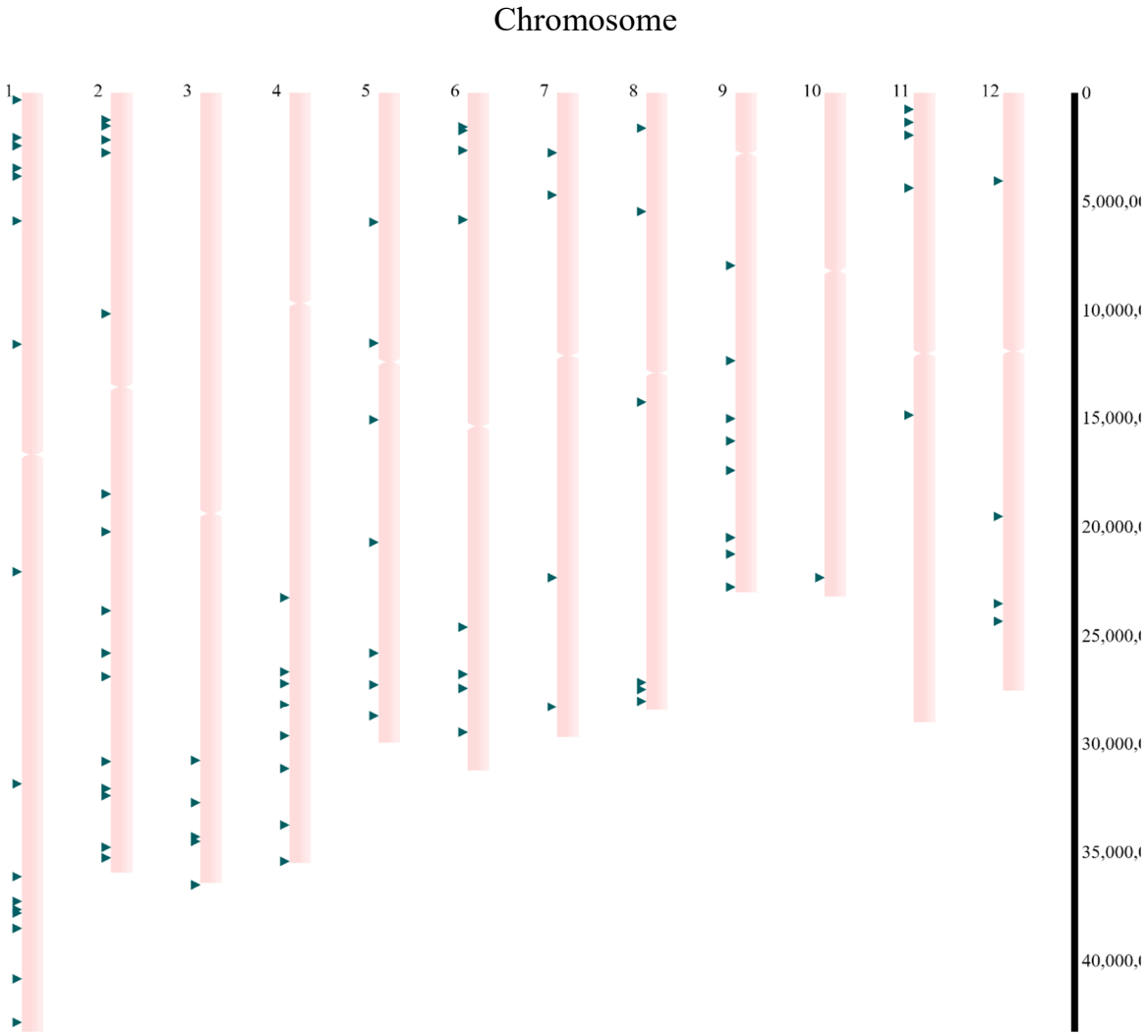


Figure 1.1. Trait genes for salinity tolerance through 12 chromosomes

Data have been replotted from the Rice SNP-Seek Database website (http://snp-seek.irri.org).

Indicates a known salinity gene on chromosomes.

Chapter 2. Profiling SNP and nucleotide diversity to characterize Mekong Delta rice landraces in Southeast Asian populations

2.1. Introduction

Recent advances in genomic analysis methods can be exploited to shed light on the genetic diversity of various crops, not only that of representative major resources, known as core collections (Gepts, 2006, Wang et al., 2014, Phan et al., 2019), but also minor resources in local collections not yet thoroughly characterized. Such evaluation of crop plant genomic structures may reveal useful resources and/or unique characteristics hidden in local collections (Hermisson & Wagner, 2004, Zhu et al., 2004, Phung et al., 2014, Penjor et al., 2016, Yousef et al., 2018, Rivera & Solis, 2019).

Genetic markers such as simple-sequence-repeat (SSR) and Restriction Fragment Length Polymorphism (RFLP) are popular in genetic diversity analysis, because the procedure is quite simple, the cost is affordable for labs with small budgets, and the analysis are quite simple. Single nucleotide polymorphism (SNP)-based genotyping is getting popular following the advance in next-generation sequencing (NGS) technologies and lower price of NGS itself (Varshney et al., 2014, Thomson 2014). The main advantages of SNP over SSR and RFLP are, 1) it is biallelic, 2) it is abundant in genome, and 3) the position in the genome is already known (Thomson 2014). All these advantages make it possible to compare results between different populations and different studies produced from various laboratories.

Single nucleotide polymorphism (SNP) genotyping methods based on next-generation sequencing and high-throughput genotyping are being applied to a variety of crops (Yang et al., 2012, Shavrukov et al., 2014, Ray & Satya, 2014, Shah et al., 2016, Chen et al., 2018). Because of their affordability, these newly improved methods can be exploited for use in small research projects with limited budgets. In particular, many currently untouched bioresource collections preserved in local institutions are now ready for characterization (Tu et al., 2007, Yamanaka et al., 2011, Mursyidin et al., 2018). Before characterization of these local resource collections, however, the most appropriate methods and genomic indicators must be determined for application in these evaluations.

Resequencing data were recently published for 3,010 rice accessions collected from 89 countries in the gene bank of the International Rice Research Institute (IRRI) (Rellosa et al., 2014, Wang et al., 2018). This rice 3K dataset contains over 29 million SNPs, 2.4 million small indels, and more than 90,000 structural variants contributing to within- and between-population variations. Population structure analysis using ADMIXTURE software classified Asian cultivated rice into nine subgroups and three admixture groups. The nine subgroups are four *indica* (XI) clusters (XI-1A from East Asia, XI-1B of improved varieties from various origins, XI-2 from Southeast Asia); three *japonica* (GJ) clusters (primarily East Asian temperate (named GJ-tmp), Southeast Asian subtropical (named GJ-sbtrp) and Southeast Asian tropical (named GJ-trp)); and two groups from South Asia: Aus, Boro and Rayada group (cA) and Sari-Basmati aromatic group (cB) (Wang et al., 2018). The public release of this polymorphism data facilitates the genomic analysis of local resources, thereby allowing the discovery of regional characteristics related to population structure, genomic diversity, and variation and contributing to variety improvement (Wang et al., 2018).

Rice produced in the Mekong Delta of Vietnam represents 54% of the total rice productivity and 90% of the total export volume of that country, which is responsible for 19% of the global rice market (Clauss et al., 2018). In the 1970s, the improved line such as IR5 started to have been introduced from IRRI to Mekong Delta, followed by dwarf varieties represented by IR8, IR36, and MTL30. Recent varieties with high productivity such as OM576, OM2451, and IR50404 have been utilized for rice production in Mekong Delta. However, it is not clear how the local rice landraces have contributed to the breeding of these modern varieties in the Mekong Delta. Although the Mekong Delta is one of the largest rice-production areas in the world, the genetic structure of rice varieties cultivated in this region has not yet been elucidated.

In this study, I focused on 99 *indica* rice accessions (81 landraces and 18 improved varieties) collected from coastal regions of 10 Mekong Delta provinces in Vietnam. These materials, a selection of *indica* accessions from the Mekong Delta Development Research Institute (MDI), are henceforth referred to as the MDI set (Fig. 2.2A & Table 2.1). Using SNP genotype data generated by double digest restriction-site associated DNA sequencing (ddRAD-seq) of the MDI set, I attempted to establish

a general procedure for the evaluation of the genomic structure of local bioresources. To achieve this objective, I used the rice 3K dataset as a reference and performed three different genomic analyses to assess the population structure, phylogenetic relationships, and diversity of the MDI set. I investigated genetic relationships between the Mekong Delta and neighboring countries and studied the genomic contribution of landraces to improved varieties in the Mekong Delta rice accession collection. Finally, I developed a simple method to profile local rice populations using low genetic diversity regions. The resulting profile is available for identification purposes and for comparison with other rice populations.

2.2. Materials and methods

2.2.1. Plant materials

Ninety-nine rice accessions, consisting of 81 local and 18 improved rice accessions, were chosen from the gene bank of the Mekong Delta Development Research Institute, Can Tho University, Vietnam (Table 2.1). The local rice accessions (MDI landraces) were collected from the Mekong Delta of Vietnam, and the 18 improved rice accessions (MDI improved varieties) were obtained from the Mekong Delta Development Research Institute. All accessions are from irrigated lowland ecosystem. My materials did not include some Mekong Delta leading varieties such as OM576, OM2514, and IR50404, although I considered that the pedigrees for the 18 improved varieties used in this study may share the similar genetic structures.

2.2.2. RAD-seq library preparation

Sterilized seeds of each variety were placed in Petri dishes and germinated in an incubator. Three days after germination, the Petri dishes were transferred into a growth chamber, and leaves were harvested for DNA extraction from 10 plants at 7–10 days after germination. DNA was extracted from approximately 100 mg of fresh young leaves using a DNeasy Plant Mini kit (Qiagen, USA) following the manufacturer's protocol. The quantity of DNA was measured using a Thermo Scientific NanoDrop 2000 spectrophotometer (Fisher Scientific, USA), and DNA quality was checked by 1%

agarose gel electrophoresis.

Each DNA sample was digested with EcoRI–BglII enzymes and ligated to RAD-seq adapter sets (Baird et al., 2008, Peterson et al., 2012). The DNA-adapter sets were pooled into a library and sequenced on a HiSeq2500 system using the 100-bp paired-end sequencing method (Illumina, USA).

2.2.3. ddRAD-seq data processing

The sequencing raw reads were filtered and sorted according to the original sample names. The sequences were trimmed to a length of 100 bp (including 5 bp of the restriction fragment plus 64 bases having a minimum quality score of 10) using Trimmomatic (Bolger et al., 2014) with the following parameters: LEADING:19, TRAILING:19, SLIDINGWINDOW:30:20, AVGQUAL:20, and MINLEN:51. The high-quality reads were mapped to the Nipponbare IRGSP1.0 *japonica* rice reference genome using Bowtie2 (Langmead & Salzberg, 2012) available in Galaxy (www.http://galaxy-mel.genome.edu.au). The reads were further filtered using Picard (http://broadinstitute.github.io/picard), and the alignments were adjusted around indels using the IndelRealigner tool provided in Genome Analysis Toolkit v2.8 (McKenna et al., 2010). SNP calling was performed with the UnifiedGenotyper tool in GATK v2.8 (DePristo et al., 2011). A total of 315,625 SNPs distributed across 12 chromosomes were identified in the 99 rice accessions (Table 2.2).

2.2.4. Phylogenetic and population structure analyses

For genetic diversity analyses, the initial SNP dataset was filtered according to the following parameters: missing call rate, 95%; minor allele frequency (MAF), 0.05; and heterozygosity rate, 0.2. The resulting dataset of 6,369 high-quality SNPs was filtered with missing call rate 100%, MAF (0.05) and Heterozygosity (0.02), 2,301 SNPs was identified. Compared with the SNPs of the 1,789 i*ndica* accessions in the 3K SNP dataset (29 million SNPs), the final combined SNP dataset consisted of 2,301 SNPs for genetic diversity analyses. PCA of the MDI set and *indica* accessions was performed in TASSEL 5.2.43 (Bradbury et al., 2007), and the results were plotted using R.

Phylogenetic analysis was carried out using the neighbor-joining method as implemented in TASSEL v5.2.43 (Bradbury et al., 2007), and visualization of the resulting tree was performed using FigTree v1.43 (http://tree.bio.ed.ac.uk/software/figtree/).

Population structure analysis was performed using ADMIXTURE v1.23 (Alexander et al., 2009, Alexander & Lange, 2011). The analysis was run with a cross-validation procedure for $K = 2$ to 9. The lowest cross-validation error was achieved at $K = 4$, which was therefore chosen as the optimal number of population partitions.

2.2.5. Nucleotide diversity analysis

Using 2,031 SNPs, I performed an analysis of the nucleotide diversity of 99 MDI rice accessions from the Mekong Delta and 412 Ind3 rice accessions from the 3K dataset (Fig. 2.1). The 412 Ind3 rice accessions are from eight Southeast Asian countries. The π-value is an index indicating nucleotide diversity at SNP site; the low π-value express the low nucleotide differences among the samples. Computation of π-values for nucleotide diversity was performed using the diversity function of TASSEL v5.2.43 (Bradbury et al., 2007) based on the sliding window option with step and window sizes of five SNPs.

2.3. Results and Discussion

2.3.1. ddRAD-seq of selected MDI rice accessions

The MDI set consisted of 81 landraces (MDI landraces) and 18 improved varieties (MDI improved varieties) housed in the gene bank of the Mekong Delta Development Research Institute, Can Tho University, Vietnam (Table 2.1). The materials were collected from 10 provinces, all located in coastal regions of the Mekong Delta in Vietnam (Fig. 2.2A & Table 2.1). The ddRAD-seq analysis was performed on all genomic DNAs of the MDI set. A total of 315,625 SNPs distributed across 12 chromosomes were identified in 99 rice varieties by ddRAD-seq. This SNP dataset was further filtered using the criteria of a 100% call rate, a 0.05 minimum allele frequency, and a 0.02 heterozygosity rate, thereby yielding 2,301 SNPs distributed across 12 chromosomes (Fig. 2.1).

Using the 2,301 SNPs, phylogenetic relationships of the 99-accession MDI set were examined by the neighbor-joining method (Fig. 2.2B). In the resulting phylogenetic tree, MDI improved varieties were clearly distinguished from MDI landraces (Fig. 2.2B). The MDI set can be divided into seven groups based on provincial origins. However, the population structure of the MDI set as determined by an ADMIXTURE analysis (Novembre, 2014) was best represented by four clusters ($K = 4$; Fig. 2.3), but no clear clustering based on provincial origin was observed. I thus assume that the MDI landrace accessions have been freely transferred among regions. Another result, namely, the finding that some accessions (MDI-42: HUYET RONG LONG AN; MDI-129: LUA DO) (Fig. 2.2B) cultivated in different regions have highly similar genotypes based on the 2301 SNPs, also supports this point of view. The observed clusters were consistent with the topology of the neighbor-joining phylogenetic tree (Fig. 2.2B & Fig. 2.3).

2.3.2. Integration of the MDI set with the 3K dataset

The 2,301 SNPs from the MDI set were analyzed along with the corresponding SNPs in the rice 3K dataset published by the IRRI (Wang et al., 2018). To reveal the relationships of the MDI accessions to those of the 3K set, I generated a phylogenetic

tree using these SNP data. As expected given the above results, MDI landraces and improved varieties were distantly located from one another in the integrated phylogenetic tree (Fig. 2.4A). The MDI landraces fell into cluster Ind3 of the 3K set, which consisted of accessions from Southeast Asian countries (Ali et al., 2011). In contrast, MDI improved varieties were found in cluster Ind1B, where most popular *indica* varieties, such as IR42 and IR64, were included (Wang et al., 2018). To elucidate the relationship between MDI landraces and Ind3 accessions from Southeast Asian countries, I conducted a principal component analysis (PCA) (Reich et al., 2008) using 81 MDI landraces and 412 Ind3 accessions derived from eight Southeast Asian countries (Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Thailand, and Vietnam) in the 3K set (Wang et al., 2018) (Fig. 2.4B & Fig. 2.5). The PCA clustered the 493 accessions into four groups (Figure 2.4B): a Vietnam+Cambodia group with a few accessions from Thailand, a Thailand+Laos group with a few accessions from Myanmar, a Philippines+Myanmar group with a few accessions from Thailand and Indonesia, and an Indonesia+Malaysia group with a few accessions from the Philippines. These results suggested that there is correlation between genetic diversity and geographical origin within the Ind3 subgroup. In fact, several previous reports also showed the correlations between the genome constitutions and geographical locations in the rice accessions distributed in Southeast Asia (Tang et al., 2010, Ali et al., 2011, Inta et al., 2016, Pusadee et al., 2017, Wang et al., 2018). Exceptionally, the group containing the accessions from the Philippines and Myanmar was geographically discontinuous (Fig. 2.5).

The PCA indicated that the MDI landraces mostly belonged to the Vietnam+Cambodia group (Fig. 2.4B), and the neighbor-joining phylogenetic analysis distributed the MDI landraces into two clusters (Fig. 2.4C). Most of the MDI landraces clustered together on a major branch of the tree, with the remaining ones incorporated into a cluster containing the Cambodian accessions (Fig. 2.4C & Fig. 2.5). The MDI landraces falling into the Cambodian group were collected from provinces close to Cambodia, such as An Giang and Kien Giang (Fig. 2.2, Fig. 2.4C & Fig. 2.5).

2.3.3. Nucleotide diversities of MDI landraces and improved varieties

When evaluating genetic resources, genetic diversity is an important factor, which can be measured by nucleotide diversity. The π-value has been utilized as an index for nucleotide diversity. The nucleotide diversity of the 99 accessions of the MDI appeared to be representative of all MDI accessions. Using the 2,301 SNPs, I separately analyzed the nucleotide diversities of landraces and improved varieties in the MDI set. To estimate the genome-wide nucleotide diversity profiles of these two groups, I performed a nucleotide diversity analysis in TASSEL v5.2.43 (Bradbury et al., 2007) by the sliding window method, with step and window sizes of five SNPs and the entire genome divided into 459 regions. The resulting π-value profiles of MDI landraces and MDI improved varieties, representing the degree of nucleotide diversity at each genomic region, were markedly different between the two groups (Fig. 2.6).

Low- and high-diversity regions were defined as those with π-values below 0.05 or over 0.45, respectively (Tables 2.3 & 2.4). In the case where the π-values used below 0.10 or over 0.40, the corresponding regions increased more than two folds of the cases with the π-values below 0.05 or over 0.45. Therefore, I selected the regions with π-values below 0.05 or over 0.45 to give more stringent thresholds. MDI landraces and improved varieties contained 15 and 18 low-diversity regions, respectively (Table 2.3), while high diversity was correspondingly indicated in 30 and 52 regions (Table 2.4).

Figure 2.7 is a graphical representation of the two nucleotide diversity profiles with different colors indicating π-value levels at each five-SNP interval throughout the genome. As shown in Fig. 2.7, low- and high-diversity regions were located in different genomic positions between the two groups. The fact that no low-diversity regions were rarely shared between MDI landraces and MDI improved varieties, implies that MDI landraces and MDI improved varieties are far to each other in terms of genetic distance, as shown in phylogenetic trees (Fig. 2.2 & Fig. 2.4A). Therefore, I assumed that these genomic regions in MDI landraces might have little contribution to genomic structures of MDI improved varieties.

2.3.4. Comparisons of genetic diversities among landraces in Southeast Asian countries

Because MDI landraces were grouped with Vietnamese and Cambodian rice accessions (Fig. 2.4C & Fig. 2.5), I assessed whether they shared similar genetic diversity profiles. Using the same 2,301 SNPs as in the above analysis, I estimated genetic diversities of 459 regions (every five SNPs) with 412 accessions from eight countries in Southeast Asia (Fig. 2.8).

A pairwise Pearson correlation analysis of genome-wide nucleotide diversity also confirmed that the profile of the MDI landraces was closest to those of accessions from Vietnam ($r = 0.69$), Cambodia ($r = 0.64$), and Thailand ($r = 0.45$); in contrast, the genome-wide nucleotide diversity profile of MDI landraces differed from that of Malaysian ($r = 0.41$) and Indonesian ($r = 0.25$) accessions (Table 2.5). These results are consistent with the groupings uncovered by the PCA (Fig. 2.4B).

The above-mentioned geographical groups discovered by the π-value analysis were correlated with locations of low and high genetic diversity in the genome (Fig. 2.9). When I plotted and compared the locations of low- and high-diversity genomic regions ($\pi < 0.05$ and $\pi > 0.45$, respectively) from each accession group from the eight Southeast Asian countries on a genomic map (Fig. 2.9), I observed a geographical correlation pattern that resembled result obtained from the PCA (Fig. 2.4B & Fig. 2.5).

Among these genomic regions, three regions of low diversity ($\pi < 0.05$) that were common to more than six populations from the eight countries and MDI landraces were found on chromosomes 1 (Chr 1), 6, and 7 (Fig. 2.9, Fig. 2.8, & Table 2.3). These three low-diversity regions were also detectable when all 493 accessions from Southeast Asian countries and MDI landraces were comprehensively examined in TASSEL v5.2.43 using the sliding window method with step and window sizes of five SNPs (Fig. 2.8). In MDI landraces, but not MDI improved varieties, all three regions were marked by low diversity (Fig. 2.7, Fig. 2.6, & Table 2.3). Consequently, these low-diversity regions may be the result of purifying selection and may include genes required for specific adaptation to Southeast Asian conditions. According to RAP-DB (https://rapdb.dna.affrc.go.jp/) and Rice SNP-Seek (http://snp-seek.irri.org/) databases, only two known genes reside in the low-diversity region of Chr 1: *OsFBX7* encoding an F-box containing protein (Hsu et al., 2004) and *ZOS1* encoding a C2H2 zinc-finger

protein (Imran et al., 2016) (Table 2.6). The known gene *RICE FLOWERING LOCUS T1* (*RFT1*) (Komiya et al., 2008) is present in the low-diversity region of Chr 6 (Table 2.6). In contrast, no known genes in these databases could be aligned with the low-diversity region of Chr 7 (Table 2.6). The genes present in these low-diversity regions may be involved in specific adaptation in Southeast Asian countries; however, *Sh4* (Martin & Busconi, 2000), *qSH1* (Konishi et al., 2006), *Sd1* (Ashikari et al., 2002), *Wx* (Hirano & Sano, 1991), *Badh2.1* (Kovach et al., 2009), and *Rc* (Furukawa et al., 2007), identified as low genetic-diversity genes in analyses of the 3K set, were not included in these three regions (Wang et al., 2018).

Aside from shared low-diversity regions, I identified seven peaks of high nucleotide diversity with a π-value greater than 0.45 that were common to more than six of the eight Southeast Asian subgroups (Fig. 2.9, Table 2.4). This high genetic diversity suggests the presence of positive selection. Although I cannot currently link the high-diversity regions to such positive selection, the fact that the high-diversity region found at 14.1–14.7 Mb on Chr 12 was shared by all Southeast Asian groups is surely not coincidental (Table 2.4).

In addition to shared low- and high-diversity regions, I respectively found four and three low- and high genetic diversity regions specific to MDI accessions (Tables 2.3 & 2.4). These regions can be used to distinguish MDI landraces from other groups collected from Southeast Asian countries. The four low-diversity regions unique to MDI landraces comprised three regions on Chr 1 and one on Chr 7 ranging from approximately 0.1–3.2 Mb, while high-diversity regions unique to MDI landraces were mapped to Chr 1, Chr 4, and Chr 10. Given that more than 1,000 genes are distributed in these regions (Table 2.3 & 2.4), I cannot presently ascertain why and how the low and high genetic diversities of these regions are specific to MDI landraces. Nevertheless, these characteristic regions can be used for MDI-landrace profiling.

2.3.5. Simple profiling for low genetic diversity

As shown in Fig. 2.9 and Fig. 2.8, I characterized the genetic diversities of MDI landraces and rice populations from eight Southeast Asian countries. I especially focused on low-diversity regions ($\pi < 0.05$) because such regions are often associated

with factors such as environmental adaptation, geography, and historical events inducing genetic drift. As shown in Fig. 2.10, I developed simple profiles of these low-diversity regions arranged in a circular layout based on the complete rice genome. These profiles, which were constructed from the 459 five-SNP interval regions (excluding intervals less than 500 kbp) of the 2,301 SNPs, allowed us to visualize the genetic diversities of the nine rice groups. The outcome of these profile comparisons was consistent with the results of our phylogenetic, principal component, and pairwise correlation analyses. These profiles of low genetic diversity in the genome can serve as fingerprints for local rice populations. The SNP information from the IRRI rice 3K set can be used to compare rice populations and determine their genetic inter-relationships.

2.4. Conclusions

As demonstrated in this study, the determination of SNP compositions of genomes of local rice populations is a powerful tool for comparing and characterizing their population structures. Our comparison of genome-wide SNPs between MDI landraces and improved varieties revealed their diverse profiles and different phylogenetic relationships. Genomic structural analyses using the 3K dataset provided by the IRRI suggested that MDI landraces belong to the Ind3 group along with other Southeast Asian landraces, whereas MDI improved varieties clustered together with the Ind1B group comprising modern improved varieties such as IR64. Our analyses revealed that the genomes of MDI landraces have contributed little to those of MDI improved varieties. Phylogenetic, principal component, and pairwise correlation analyses demonstrated the different genetic properties of the 412 Ind3 accessions from eight Southeast Asian countries in the 3K set and sorted them into four groups possibly related to geography. The MDI landraces were closely related to the Vietnam+Cambodia group but were far from the Indonesia+Malaysia group. These relationships were also clearly reflected in the pattern of nucleotide (genetic) diversity. Low and high π-value regions in each of the eight populations were mapped onto the chromosomes, and those common to most of the eight populations or that were population specific were graphically highlighted. Low π-value regions are particularly useful as unique genomic signatures of local accessions. I thus propose that simple

profiling of a local rice population using low π-value regions can be used for identification purposes.

Table 2.1. Accession coding number, local name, origin, and species subgroup of Mekong rice of Vietnam

Accession code	accession name	District	Province	Species subgroup
MDI-1	BO LIEP 2		Ca Mau	Landrace
MDI-2	BA BONG MAN		Ca Mau	Landrace
MDI-3	LUN PHEN	Thoi Binh	Ca Mau	Landrace
MDI-4	LUN DO	U Minh	Ca Mau	Landrace
MDI-5	LUN PHEN HAT NHO	Thoi Binh	Ca Mau	Landrace
MDI-6	LUN MAN	Tran Van Thoi	Ca Mau	Landrace
MDI-7	LUN VANG	Tran Van Thoi	Ca Mau	Landrace
MDI-8	LUN PHET	U Minh	Ca Mau	Landrace
MDI-9	LUN HEN	Thoi Binh	Ca Mau	Landrace
MDI-11	LUN CAN DO	U Minh	Ca Mau	Landrace
MDI-12	LUN CAN TRANG	Tran Van Thoi	Ca Mau	Landrace
MDI-13	TRANG PHIEU	Thoi Binh	Ca Mau	Landrace
MDI-14	MOT BUI LUN	Thoi Binh	Ca Mau	Landrace
MDI-15	MOT BUI LUN CA MAU		Ca Mau	Landrace
MDI-16	MOT BUI TRANG	Thoi Binh	Ca Mau	Landrace
MDI-17	NANG QUOT BIEN	U Minh	Ca Mau	Landrace
MDI-18	NEP SUA	Tran Van Thoi	Ca Mau	Landrace
MDI-19	SOI LUN	Tran Van Thoi	Ca Mau	Landrace
MDI-21	DOC PHUNG	Binh Dai	Ben Tre	Landrace
MDI-22	TRA LONG 2		Ca Mau	Landrace
MDI-23	BA BUI 2		Ca Mau	Landrace
MDI-24	NANG QUOT BIEN 1	Tran Van Thoi	Ca Mau	Landrace
MDI-25	MOT BUI LUN 2	Thoi Binh	Ca Mau	Landrace
MDI-26	MONG CHIM DEN		Ca Mau	Landrace
MDI-27	NAM TAI 1	Thoi Binh	Ca Mau	Landrace
MDI-28	MONG CHIM ROI 3	Thoi Binh	Ca Mau	Landrace
MDI-29	BA BUI LUN	U Minh	Ca Mau	Landrace
MDI-30	MOT BUI 5	Tran Van Thoi	Ca Mau	Landrace
MDI-31	MOT BUI DO CAO CA MAU		Ca Mau	Landrace
MDI-41	THOM MUA 1		Ca Mau	Landrace
MDI-42	HUYET RONG LONG AN		Long An	Landrace
MDI-43	TAI NGUYEN SUA	Tran Van Thoi	Ca Mau	Landrace
MDI-44	MTL110		Ben Tre	Improved rice

Accession code	accession name	District	Province	Species subgroup
MDI-49	MTL944		Can Tho	Improved rice
MDI-50	MTL560		Can Tho	Improved rice
MDI-52	MTL480		Can Tho	Improved rice
MDI-53	MTL547		Can Tho	Improved rice
MDI-54	MTL939		Can Tho	Improved rice
MDI-55	MTL934		Can Tho	Improved rice
MDI-56	MTL943		Can Tho	Improved rice
MDI-57	MTL926		Can Tho	Improved rice
MDI-58	MTL941		Can Tho	Improved rice
MDI-59	MTL942		Can Tho	Improved rice
MDI-60	MTL946		Can Tho	Improved rice
MDI-61	MTL938		Can Tho	Improved rice
MDI-62	MTL945		Can Tho	Improved rice
MDI-63	MTL936		Can Tho	Improved rice
MDI-64	MTL930		Can Tho	Improved rice
MDI-65	MTL940		Can Tho	Improved rice
MDI-66	MTL372		Can Tho	Improved rice
MDI-67	Ba bui 1	Ba Tri	Ben Tre	Landrace
MDI-68	BA MUOI MUA	Binh Minh	Vinh Long	Landrace
MDI-69	BA TUC 1		Ca Mau	Landrace
MDI-70	BAY TAN	Cai Nuoc	Ca Mau	Landrace
MDI-71	BONG VANG		Kien Giang	Landrace
MDI-72	CHAU HANG VO		Kien Giang	Landrace
MDI-75	CHUM RUOT MUON	Tran Van Thoi	Ca Mau	Landrace
MDI-77	CHUOI	My Tho	Tien Giang	Landrace
MDI-82	LUA CHUOI	Ha Tien	Kien Giang	Landrace
MDI-83	NANG DUT 1	Cai Nuoc	Ca Mau	Landrace
MDI-85	NANG HUONG 1		Long An	Landrace
MDI-86	NANG HUONG TRON	Binh Minh	Vinh Long	Landrace
MDI-87	NANG KE		An Giang	Landrace
MDI-88	NANG KEO BA TU		Tra Vinh	Landrace
MDI-89	NANG KEO XIEM 2	Cau Ngang	Tra Vinh	Landrace
MDI-90	NANG NUOOL 3		An Giang	Landrace
MDI-91	NANG PHET 3		Ca Mau	Landrace
MDI-92	NANG QUOT 1	Dam Doi	Ca Mau	Landrace

Accession code	accession name	District	Province	Species subgroup
MDI-93	NANG QUOT 2		Ca Mau	Landrace
MDI-95	NEP DAI LOAN 2	Dam Doi	Ca Mau	Landrace
MDI-96	NEP MO 3	Dam Doi	Ca Mau	Landrace
MDI-97	NEP RUOI SOM 1	Ba Tri	Ben Tre	Landrace
MDI-98	NEP TAM SAC		Long An	Landrace
MDI-99	NEP THAN NONG 1		Ca Mau	Landrace
MDI-100	NEP TRANG 1	Tran Van Thoi	Ca Mau	Landrace
MDI-101	SAI GON		Hau Giang	Landrace
MDI-102	SOI DO		Tra Vinh	Landrace
MDI-104	TRANG CHI SUOI		Ca Mau	Landrace
MDI-105	TRANG CHUM LUA		Tra Vinh	Landrace
MDI-106	TRANG LUN 3	Tra Cu	Tra Vinh	Landrace
MDI-108	TRAU TRON		Hau Giang	Landrace
MDI-109	TRUNG KIEN		Kien Giang	Landrace
MDI-111	VOI		Ben Tre	Landrace
MDI-112	HAI HOANH	Tra Cu	Tra Vinh	Landrace
MDI-118	THAN NONG DO	Ha Tien	Kien Giang	Landrace
MDI-119	MAKARI	Ha Tien	Kien Giang	Landrace
MDI-120	BONG GUNG	Kien Luong	Kien Giang	Landrace
MDI-121	NANG NONLUC	Ha Tien	Kien Giang	Landrace
MDI-122	CHIM ROI	Kien Luong	Kien Giang	Landrace
MDI-123	NEP AO GIA	Ha Tien	Kien Giang	Landrace
MDI-124	NANG COI	HaTien	Kien Giang	Landrace
MDI-125	NHO HUONG	Thanh Phu	Ben Tre	Landrace
MDI-126	NANG THOM CHO DAO	Can Duoc	Long An	Landrace
MDI-127	TAI NGUYEN	Tan Tru	Long An	Landrace
MDI-128	NEP CHANHHOL	Tinh Bien	An Giang	Landrace
MDI-129	LUA DO	Tra Cu	Tra Vinh	Landrace
MDI-130	NANG TET	Tra Cu	Tra Vinh	Landrace
MDI-131	NANG KEO	Thanh Phu	Ben Tre	Landrace
MDI-133	CHO BIEN	Ha Tien	Kien Giang	Landrace

Table 2.2. SNPs information of 99 MDI rice accessions

No.	Chromosome	SNPs number
1	1	37,136
2	2	32,140
3	3	33,089
4	4	27,105
5	5	24,296
6	6	26,538
7	7	24,212
8	8	23,928
9	9	19,995
10	10	20,085
11	11	25,101
12	12	22,088
Total		**315,625**

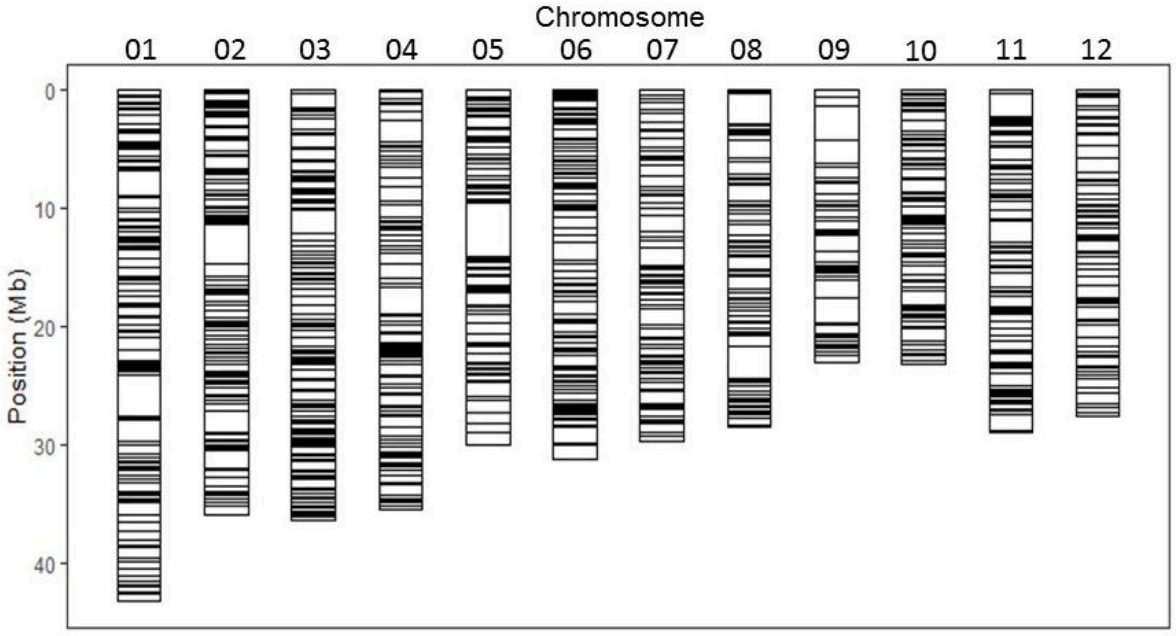


Figure 2.1. The distribution of 2,301 SNPs on the 12 rice chromosomes

Horizontal line was the position of SNPs on 12 chromosomes

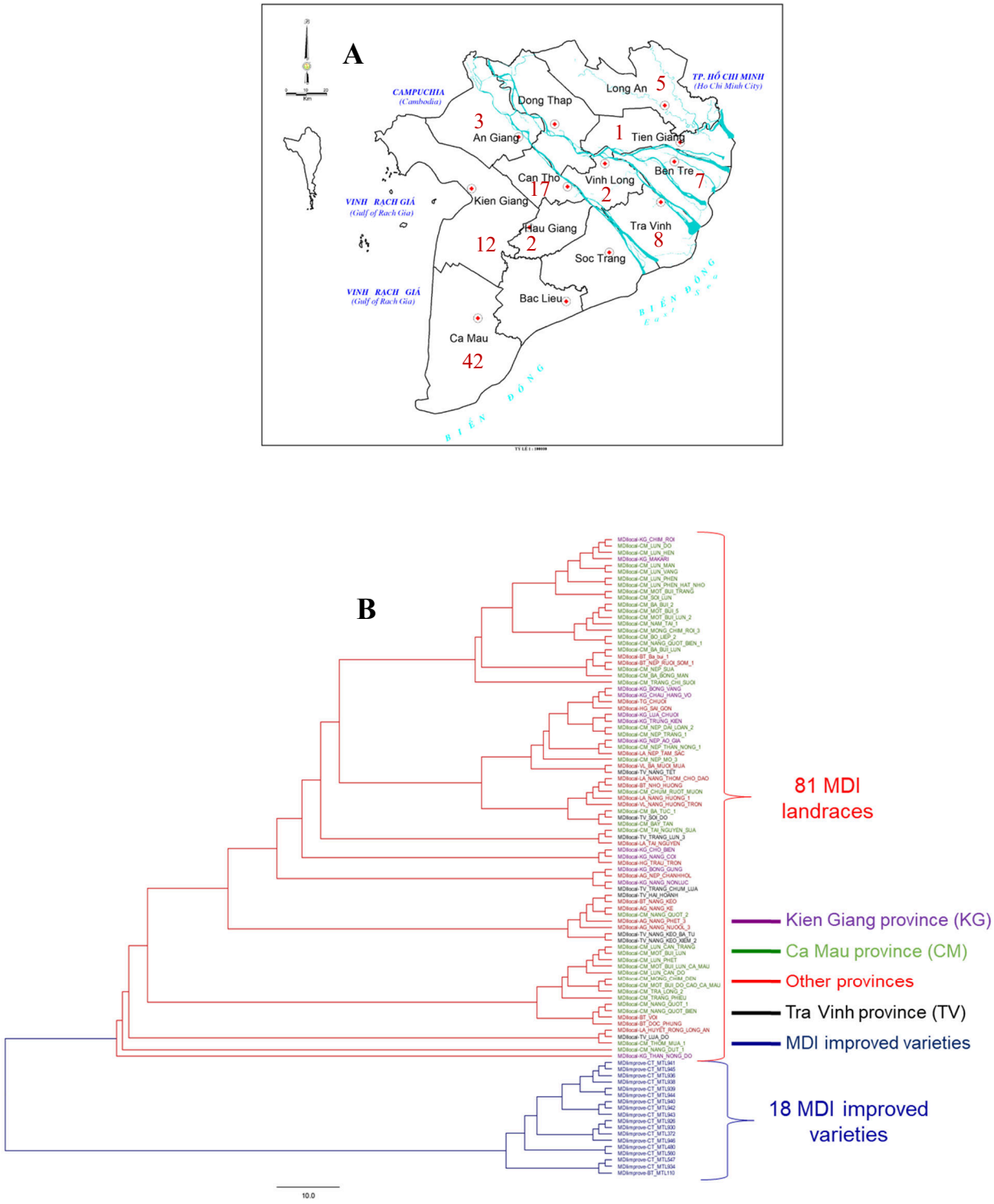


Figure 2.2. Distribution and phylogenetic relationships of 99 MDI rice accessions.

A. Distribution of 99 rice accessions from the Mekong Delta of Vietnam. Rice accessions were collected from 10 Mekong Delta provinces. The number of selected rice accessions from each province is displayed on the map. B. Phylogenetic tree based on 2,301 SNPs from the 99 MDI rice accessions. The neighbor-joining tree was generated using TASSEL v5.2.43. The red star indicates accessions with different names and locations that have the same genotype.

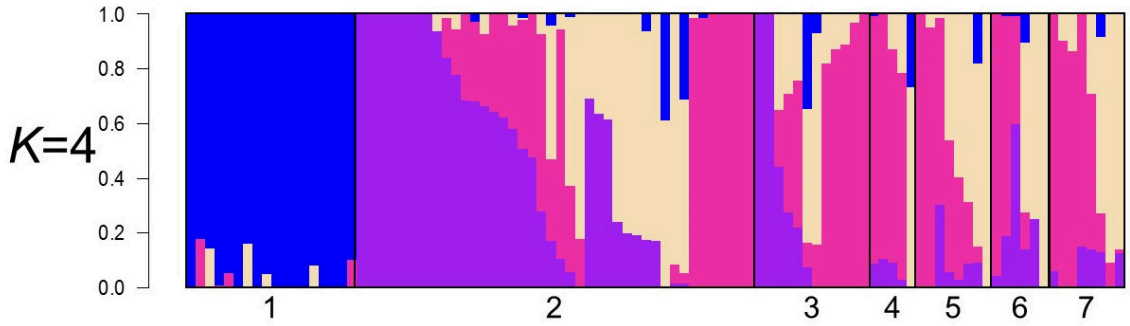


Figure 2.3. Population structure based on an ADMIXTURE analysis at *K = 4*.

Individuals are represented by vertical bars shaded in proportion to their estimated ancestry within each cluster. No clear clustering based on geographical origin was observed. All improved varieties (#1) were grouped into one cluster (blue); the remaining groups were Ca Mau (# 2), Kien Giang (#3), Long An (#4), Tra Vinh (#5), Ben Tre (#6), and other provinces (# 7).

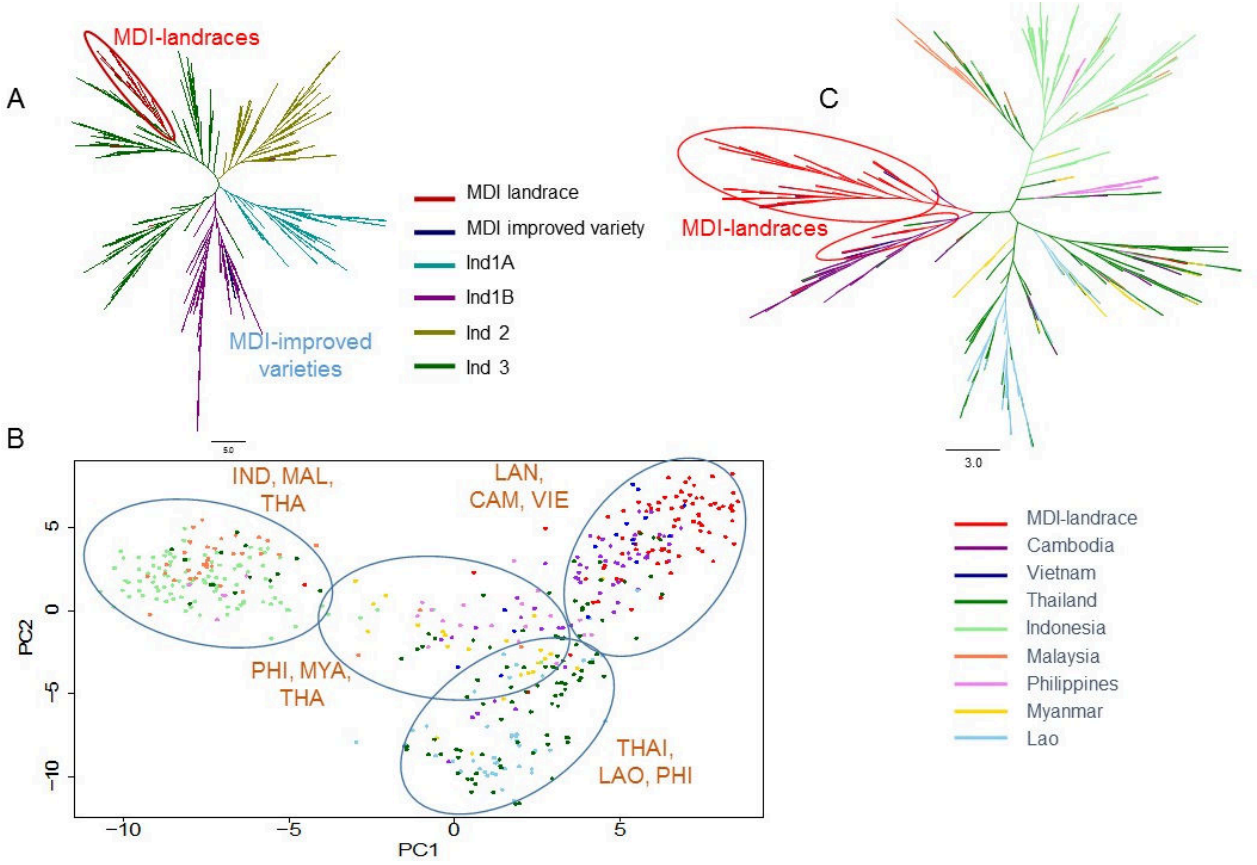


Figure 2.4. Genetic relationships of MDI rice accessions and rice accessions from the 3K rice genome dataset.

A. Phylogenetic tree of 1,174 indica accessions belonging to the 3K rice genome dataset and 99 MDI rice accessions. The 3K rice genome accessions consisted of 209 Ind1A, 205 Ind1B, 285 Ind2, and 475 Ind3 accessions, whereas the MDI rice dataset contained 81 landraces and 18 improved varieties. The tree was generated by the neighbor-joining method using 2,301 SNPs common to both the 3K rice genome dataset and the MDI dataset. B. Results of principal component analysis of 412 Ind3 accessions from Southeast Asian countries and 81 MDI landrace accessions. The number of rice accessions from each dataset or country is given in parentheses as follows: MDI landrace (81), Thailand (107), Indonesia (103), Cambodia (55), Laos (47), Malaysia (37), the Philippines (24), Myanmar (23), and Vietnam (16). C. Neighbor-joining phylogenetic tree of 412 Southeast Asian Ind3 accessions and 81 MDI landraces generated from 2,301 SNPs common to both 3K rice genome and MDI datasets.

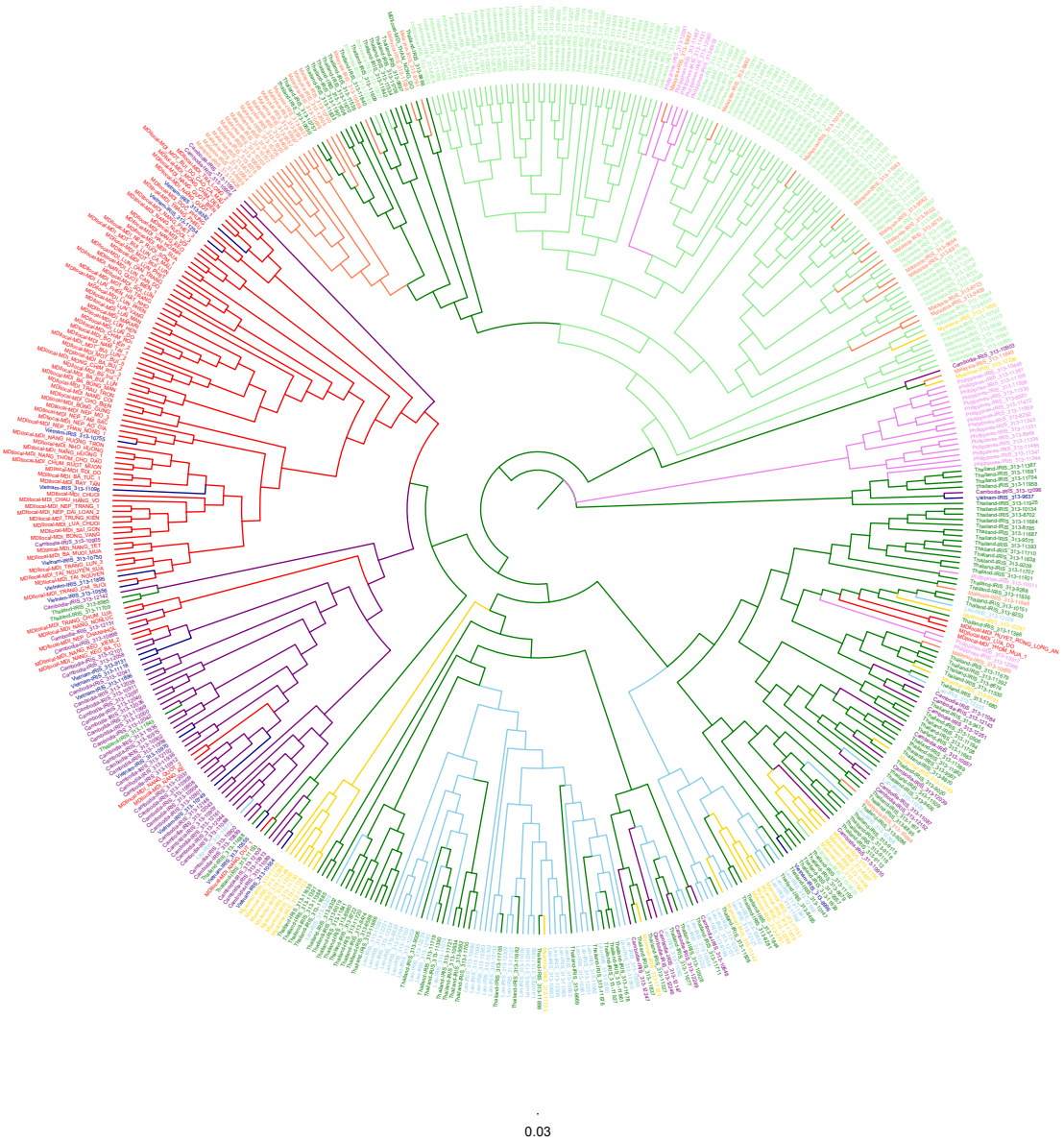
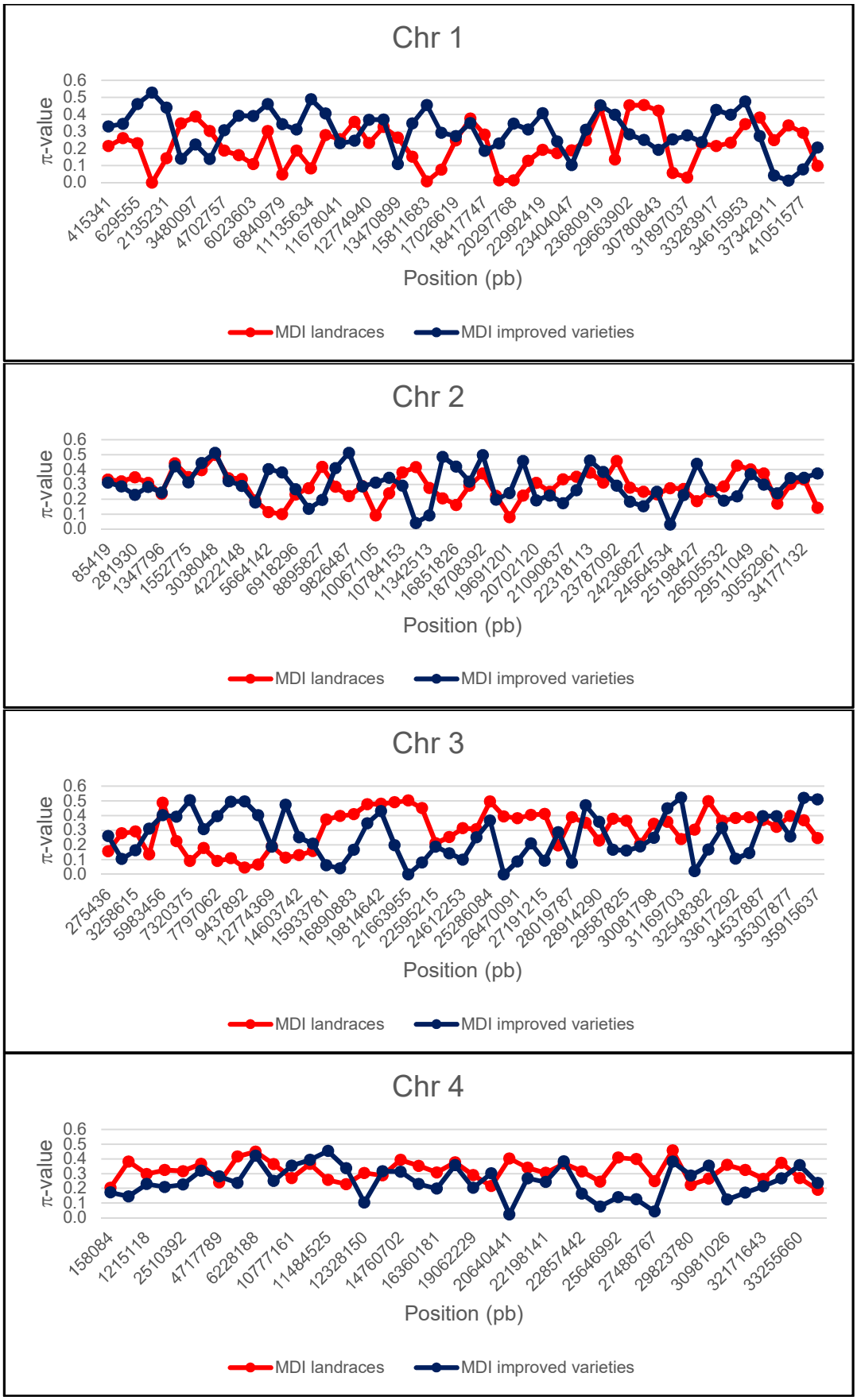
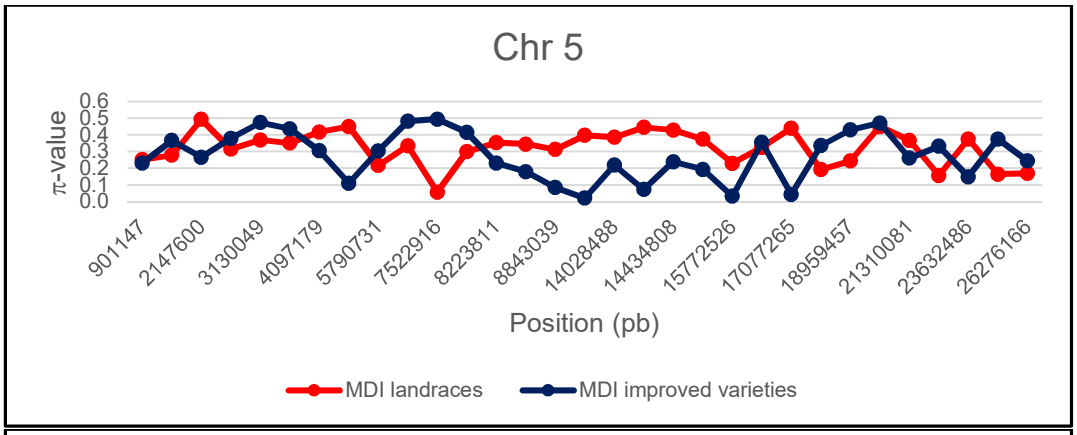
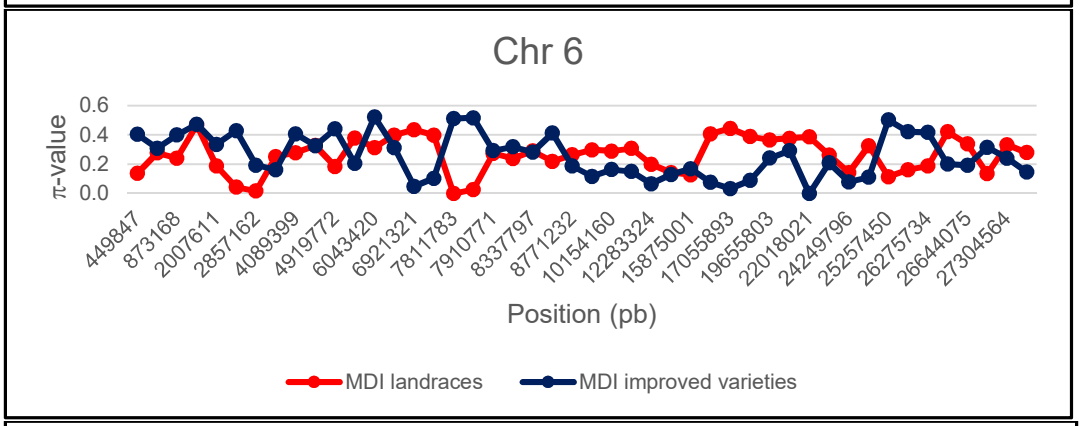


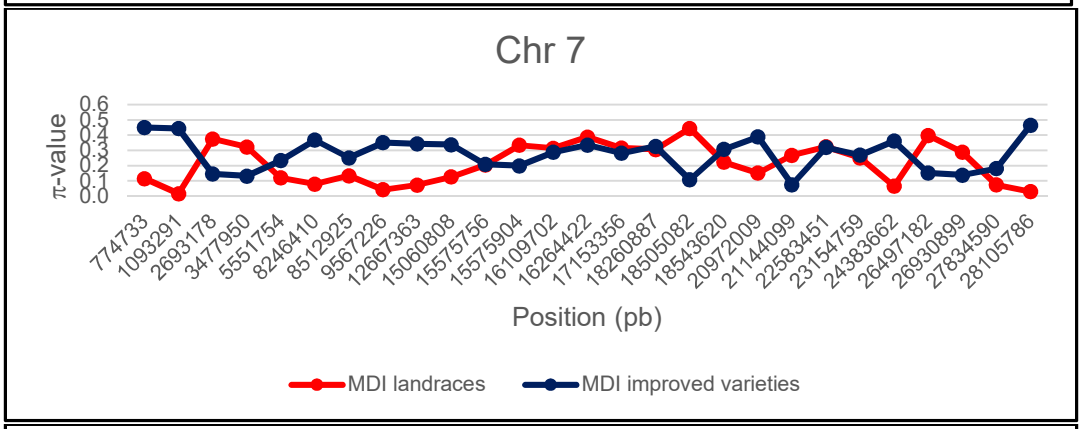
Figure 2.5 Phylogenetic tree of 412 Southeast Asian ind3 accessions and 81 MDI landraces.

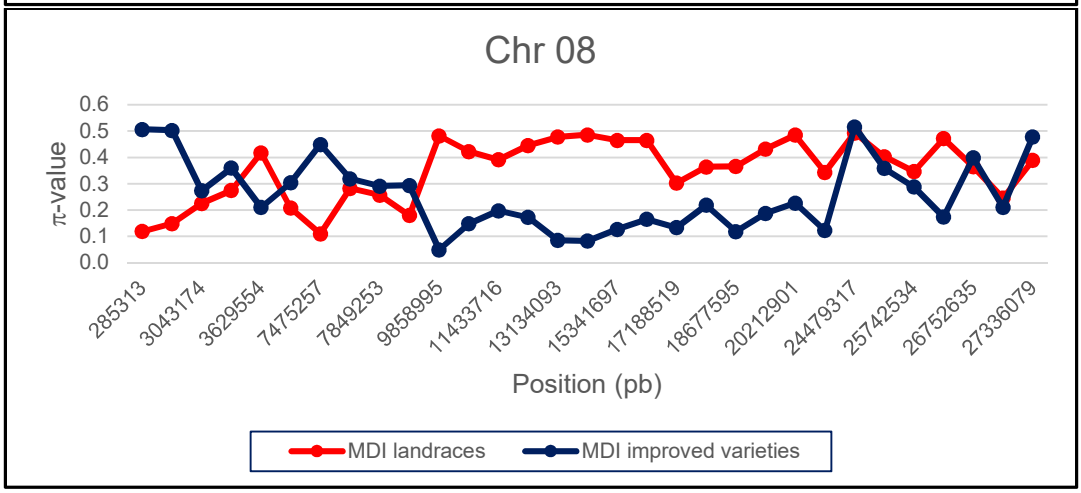
The tree was generated by the neighbor-joining method using 2,301 SNPs common to both the 3K rice genome dataset.


Chr 1
π-value
0.6
0.5
0.4
0.3
0.2
0.1
0.0
415341
629555
2135231
3480097
4702757
6023603
6840979
11135634
11678041
12774940
13470899
15811683
17026619
18417747
20297768
22992419
23404047
23680919
29663902
30780843
31897037
33283917
34615953
37342911
41051577
Position (pb)
MDI landraces
MDI improved varieties
Chr 2
π-value
0.6
0.5
0.4
0.3
0.2
0.1
0.0
85419
281930
1347796
1552775
3038048
4222148
5664142
6918296
8895827
9826487
10067105
10784153
11342513
16851826
18708392
19691201
20702120
21090837
22318113
23787092
24236827
24564534
25198427
26505532
29511049
30552961
34177132
Position (pb)
MDI landraces
MDI improved varieties
Chr 3
π-value
0.6
0.5
0.4
0.3
0.2
0.1
0.0
275436
3258615
5983456
7320375
7797062
9437892
12774369
14603742
15933781
16890883
19814642
21663955
22595215
24612253
25286084
26470091
27191215
28019787
28914290
29587825
30081798
31169703
32548382
33617292
34537887
35307877
35915637
Position (pb)
MDI landraces
MDI improved varieties
Chr 4
π-value
0.6
0.5
0.4
0.3
0.2
0.1
0.0
158084
1215118
2510392
4717789
6228188
10777161
11484525
12328150
14760702
16360181
19062229
20640441
22198141
22857442
25646992
27488767
29823780
30981026
32171643
33255660
Position (pb)
MDI landraces
MDI improved varieties


Chr 5
π-value
0.6
0.5
0.4
0.3
0.2
0.1
0.0
901147
2147600
3130049
4097179
5790731
7522916
8223811
8843039
14028488
14434808
15772526
17077265
18959457
21310081
23632486
26276166
Position (pb)
MDI landraces
MDI improved varieties


Chr 6
π-value
0.6
0.4
0.2
0.0
449847
873168
2007611
2857162
4089399
4919772
6043420
6921321
7811783
7910771
8337797
8771232
10154160
12283324
15875001
17055893
19655803
22018821
24249796
25257450
26275734
26644075
27304564
Position (pb)
MDI landraces
MDI improved varieties


Chr 7
π-value
0.6
0.5
0.4
0.3
0.2
0.1
0.0
774733
1093291
2693178
3477950
5551754
8246410
8512925
9567226
12667363
15060808
15575756
15575904
16109702
16264422
17153356
18260887
18505082
18543620
20972009
21144099
22583451
23154759
24383662
26497182
26930899
27834590
28105786
Position (pb)
MDI landraces
MDI improved varieties


Chr 08
π-value
0.6
0.5
0.4
0.3
0.2
0.1
0.0
285313
3043174
3629554
7475257
7849253
9858995
11433716
13134093
15341697
17188519
18677595
20212901
24479317
25742534
26752635
27336079
Position (pb)
MDI landraces
MDI improved varieties

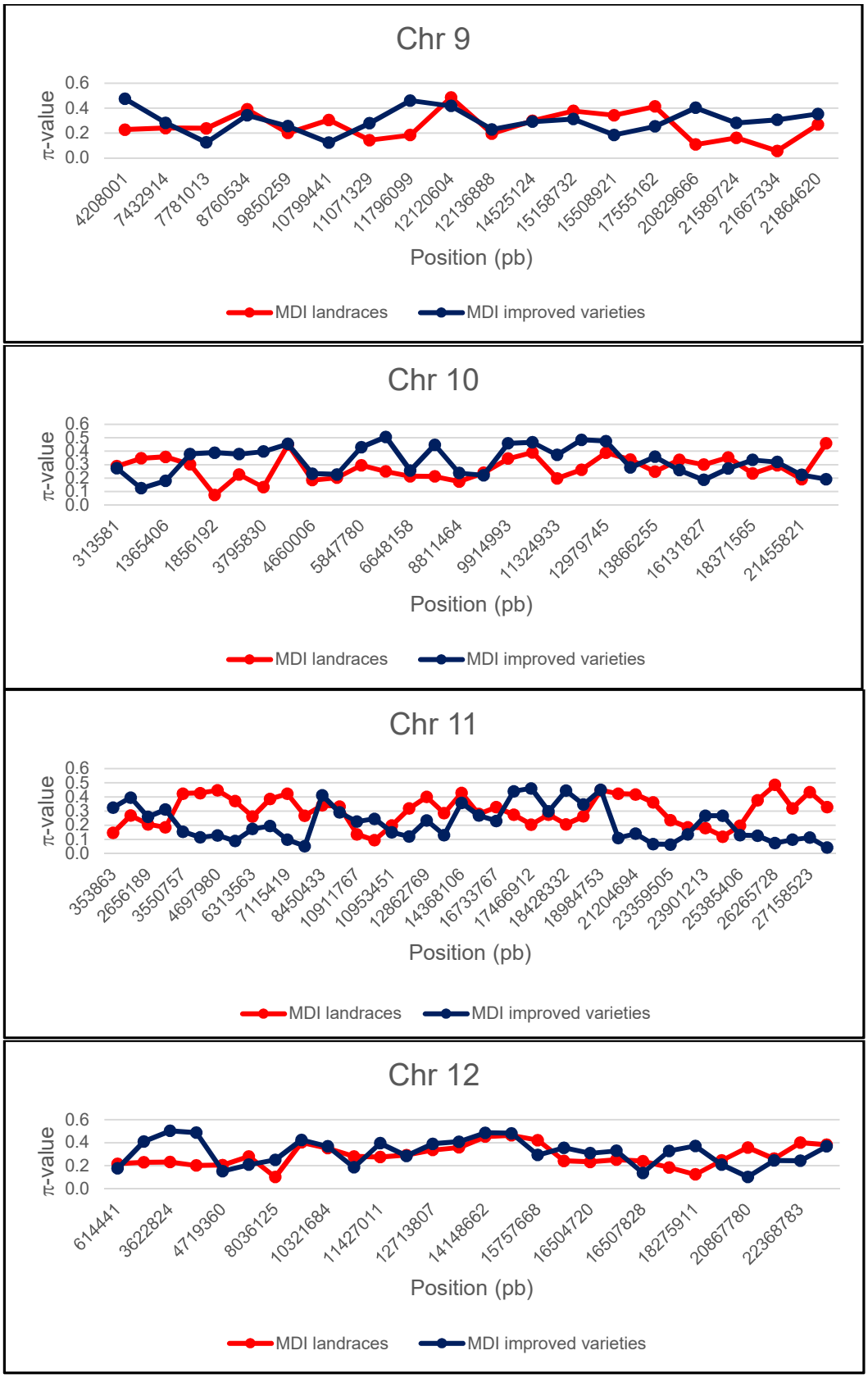


Figure 2.6. The π-values distribution of 99 MDI rice accessions through overall 459 chromosomal regions. *The distribution is separately displayed in each of the 12 chromosomes. All π-values were calculated by the sliding window method in TASSEL v5.2.43 using the diversity tool with step and window sizes of five SNPs across 2,301 SNPs. X and Y axis indicate the nucleotide positions of SNPs in the chromosomes and π-values, respectively.*

Table 2.3. Low nucleotide diversity position of Ind3 from eight Southeast Asian countries and MDI rice accessions

Chr	Start Chr Position	End Chr Position	MDI-landraces	MDI-impved varieties	Vietnam	Cambodia	Thailand	Philippines	Lao	Indonesia	Malaysia	Myanmar	**Groups of population**
1	1732588	1732622	0										**1**
1	6840979	10067618	0.04781										**1**
1	15811683	15904518	0.00741		0	0.0035	0	0		0.00189	0		**7**
1	15904562	16377268			0.02417	0.02882		0.03442			0.04535		**4**
1	19153121	20297735	0.01222										**1**
1	20297768	20996888	0.01225		0.02542	0.01096							**3**
1	29663902	29975472			0								**1**
1	31897037	31990979	0.02985										**1**
1	37342911	38752458		0.04183									**1**
1	38752562	41040618		0.01111									**1**
2	5114607	5583172									0.04384		**1**
2	11225029	11342496		0.04183									**1**
2	24236827	24268439										0.01779	**1**
2	24564534	24774053		0.03203									**1**
2	35200080	275366	0.04083										**1**
3	7320375	7349686			0.04833			0			0.02643		**3**
3	9437892	10007036	0.04602		0.02583								**2**
3	16305784	16890773		0.04118									**1**
3	21663955	22209289		0									**1**
3	24620384	24620442						0					**1**
3	25460747	26470079		0									**1**
3	27191215	27543456									0.02988		**1**
3	29854325	30047290								0.02848	0		**2**
3	31169703	31473739									0.02102		**1**
3	32174427	32548335		0.02222									**1**
4	11746267	12328128								0.04101			**1**
4	12328150	13231660						0		0	0	0.04506	**4**
4	14760702	14760826								0.00386	0.01351		**2**

Chr	Start Chr Position	End Chr Position	MDI-landraces	MDI-impved varieties	Vietnam	Cambodia	Thailand	Philippines	Lao	Indonesia	Malaysia	Myanmar	**Groups of population**
4	20640441	21530002		0.02222									**1**
4	27488767	29235162		0.04183									**1**
4	34348909	34683804							0.01272				**1**
5	7522916	7534349			0				0.04165				**2**
5	9494944	9495122		0.02222									**1**
5	15772526	16760981		0.03333									**1**
5	17077265	18266571		0.04444									**1**
5	26276166	28251689							0.02549				**1**
5	28928628	279970			0.025	0.00731	0.00557		0.02139		0.03964		**5**
6	2696934	2857107	0.04296								0.02162		**2**
6	2857162	2939468	0.01716		0	0.00364	0.00933		0.0296	0.02311	0.04809	0.03478	**8**
6	4089399	4197261							0.02109				**1**
6	6921321	7158775		0.04902									**1**
6	7811783	7848380	0		0	0.04706							**3**
6	7848385	7910673	0.02759		0	0.01827						0.02895	**4**
6	15875001	16573140									0.04711		**1**
6	17055893	17890100		0.03333									**1**
6	22018021	23336727		0									**1**
6	24948054	24948214			0.01167								**1**
6	25257450	25340359			0								**1**
7	1093291	2007499	0.01719		0	0.0383							**3**
7	5551754	7335367										0.04308	**1**
7	8512925	8913524										0.04318	**1**
7	9567226	12431967	0.04377		0.025	0.00727	0.03136	0		0.01736	0.00541		**7**
7	12667363	15057124			0.00667	0.04936				0.03729	0.01224		**4**
7	18543620	20228883									0.042		**1**
7	28105786	28908337	0.03108										**1**
8	3043174	3378446					0.03805		0.00842				**2**
8	3378538	3629549							0.03404				**1**
8	7999376	9633753							0.04898				**1**

Chr	Start Chr Position	End Chr Position	MDI-landraces	MDI-impved varieties	Vietnam	Cambodia	Thailand	Philippines	Lao	Indonesia	Malaysia	Myanmar	**Groups of population**
8	9858995	10494049		0.04935									**1**
8	19731626	20188425							0.03367				**1**
9	21667334	21773712			0.04667		0.02585		0.02555				**3**
10	1856192	1856297				0.03663	0.03417			0.01933	0		**4**
10	6648158	7544903									0		**1**
10	7544980	8811441									0.03238		**1**
11	10953051	10953365				0.04983		0.02428					**2**
11	27207479	27492897		0.04444									**1**
12	3827769	3827885			0					0.00195			**2**
12	8036125	8823074						0.01667	0.04385				**2**
	Total		**15**	**18**	**18**	**12**	**7**	**8**	**12**	**10**	**19**	**6**	

Note: Low nucleotide diversity (π-value < 0.05); totally 493 accessions consist of 81 MDI landraces, 412 rice accessions from 8 countries; specific low: only low diversity in MDI landrace accessions; common low: there were at least 7 out of 9 country groups, which were low diversity.

Table 2.4. High nucleotide diversity position of Ind3 from eight Southeast Asian countries and MDI rice accessions

Chr	Start Chr Position	End Chr Position	MDI-landraces	MDI-impved varieties	Vietnam	Cambodia	Thailand	Philippines	Lao	Indonesia	Malaysia	Myanmar	**Groups of population**
1	415341	415487						0.46377				0.45534	**2**
1	629555	1522268		0.46209						0.49803	0.48754		**3**
1	1732588	1732622		0.52941									**1**
1	5567340	6023592							0.4679	0.50144			**2**
1	6494349	6794643		0.46209									**1**
1	11135634	11135729		0.48954									**1**
1	11678041	11973292			0.45313		0.47029	0.4942		0.4724	0.46291	0.53202	**6**
1	11973354	12750071				0.46562							**1**
1	12923023	13359800									0.45968		**1**
1	13470899	14311547						0.46377					**1**
1	15811683	15904518		0.4549									**1**
1	23404047	23466140										0.51133	**1**
1	23680919	23791463		0.45425									**1**
1	29663902	29975472	0.4534										**1**
1	29975496	30759953	0.45519										**1**
1	30780843	31369604										0.47579	**1**
1	34615953	34862018		0.47647				0.49855					**2**
1	34862054	37323617								0.4636			**1**
1	41051577	41867114						0.46594					**1**
1	42616776	85401						0.51159			0.50435		**2**
2	1481244	1552772			0.47917	0.4865	0.47849	0.51386		0.45831	0.47267		**6**
2	1552775	2018938										0.49493	**1**
2	2136297	3038028						0.47391					**1**
2	3038048	3050716	0.49978	0.51242	0.5075		0.48503	0.49275	0.49084	0.46732	0.51261		**8**
2	9826487	9826527		0.51307	0.49417	0.47481	0.48584						**4**
2	10067105	10313147									0.45075		**1**
2	16012297	16851817		0.48627									**1**
2	18708392	19560936		0.49804									**1**

Chr	Start Chr Position	End Chr Position	MDI-landraces	MDI-impved varieties	Vietnam	Cambodia	Thailand	Philippines	Lao	Indonesia	Malaysia	Myanmar	**Groups of population**
2	19560993	19691196			0.50167					0.46935			**2**
2	19924630	20383553		0.45817									**1**
2	20702120	21028794					0.48391						**1**
2	21028826	21030158					0.49033		0.49584				**2**
2	22318113	22877691		0.46144									**1**
2	23787092	23982182	0.45744			0.45372							**2**
2	29511049	29619186										0.48933	**1**
2	33578837	34164910									0.47071		**1**
2	34177132	34269184						0.4808					**1**
3	3853849	4948986					0.47984					0.46469	**2**
3	5983456	5998011	0.48889			0.50079	0.49339	0.48333		0.50155			**5**
3	7320375	7349686		0.50588									**1**
3	9207810	9384873		0.49542				0.48986					**2**
3	9437892	10007036		0.49739				0.48043					**2**
3	13988304	14603612		0.47451									**1**
3	14603742	14677220									0.46014		**1**
3	16890883	18879288										0.49723	**1**
3	18879291	19814541	0.47744		0.49896	0.46896		0.46377	0.47017			0.51996	**6**
3	19814642	20963393	0.48071		0.5	0.45961		0.46585	0.46531			0.5191	**6**
3	20963412	21663855	0.49133		0.51083	0.47285		0.45072	0.48205			0.52022	**6**
3	21663955	22209289	0.50299		0.51167	0.49909	0.47395		0.48571			0.49634	**6**
3	22248728	22581548	0.45093			0.50088	0.4849					0.46897	**4**
3	25286084	25459768	0.49623		0.46736	0.50187		0.4971				0.48969	**5**
3	26470091	26681986					0.46089						**1**
3	28656654	28914254		0.46993		0.46448		0.45127					**3**
3	30874284	31169650					0.45005	0.46341				0.45099	**3**
3	31169703	31473739		0.52353				0.47246	0.45319			0.48221	**4**
3	32548382	32581892	0.49926					0.51377					**2**
3	34537887	34898837						0.46196	0.50879	0.48231			**3**

Chr	Start Chr Position	End Chr Position	MDI-landraces	MDI-impved varieties	Vietnam	Cambodia	Thailand	Philippines	Lao	Indonesia	Malaysia	Myanmar	**Groups of population**
3	35307877	35837909					0.46829					0.45346	**2**
3	35901640	35915546		0.52157			0.4989				0.48348	0.48379	**4**
3	35915637	36101974		0.51111			0.46241		0.45116				**3**
3	36102069	89823		0.48562									**1**
4	698689	1215043			0.475								**1**
4	6228188	6478334	0.45015						0.49648				**2**
4	11484525	11650866		0.4549									**1**
4	16607874	19010311				0.48907		0.46159		0.45379		0.51117	**4**
4	29235181	29823746	0.45941										**1**
4	30573131	30789687						0.48134					**1**
4	32171643	32171781			0.45167								**1**
4	33255660	33430081								0.4777	0.45313		**2**
5	1476072	2101578								0.46503			**1**
5	2147600	2249528	0.49306		0.49736		0.49111					0.48379	**4**
5	3130049	3874317		0.47451					0.49534	0.47797			**3**
5	3874356	4097133				0.50912		0.51027	0.49919	0.49793	0.51336	0.52194	**6**
5	4097179	4815791			0.47								**1**
5	4815845	5507776			0.52333			0.45942			0.47245	0.4917	**4**
5	6654199	7522914		0.4817						0.47316	0.46881	0.48205	**4**
5	7522916	7534349		0.49477							0.46637		**2**
5	15187459	15772525			0.51833	0.48882							**2**
5	17077265	18266571			0.45313					0.49129	0.46441		**3**
5	20555311	21310073	0.45012	0.47124	0.45833			0.47681		0.47133	0.47661	0.46245	**7**
6	1726063	2005912	0.46904	0.4732		0.47769	0.47863	0.5029					**5**
6	4919772	5182054						0.47029					**1**
6	6043420	6303564		0.52353			0.50389	0.47138		0.49651		0.50336	**5**
6	6303577	6723750						0.4529					**1**
6	7158778	7811635										0.4666	**1**
6	7811783	7848380		0.51176							0.50976		**2**

Chr	Start Chr Position	End Chr Position	MDI-landraces	MDI-impved varieties	Vietnam	Cambodia	Thailand	Philippines	Lao	Indonesia	Malaysia	Myanmar	**Groups of population**
6	7848385	7910673		0.51699									**1**
6	8337644	8337791						0.48406	0.50968			0.50138	**3**
6	8353306	8771191						0.50072	0.48086			0.4941	**3**
6	10691000	11618394			0.485								**1**
6	16912709	16984804			0.49125	0.469							**2**
6	19655803	20959464					0.49362					0.45909	**2**
6	25257450	25340359		0.50261									**1**
6	25340364	26263837									0.46813		**1**
6	27879718	28580644				0.46189		0.49384					**2**
7	774733	1053247		0.45033									**1**
7	15060808	15210706						0.47428					**1**
7	15575756	15575887					0.49382				0.51065	0.4863	**3**
7	15575904	16068606					0.48814						**1**
7	16264422	16656924			0.4725								**1**
7	20972009	21132761					0.45594			0.46393	0.48565		**3**
7	21144099	22388616				0.47508							**1**
7	28105786	28908337		0.46536									**1**
8	285313	2905371		0.50654									**1**
8	2905388	2905470		0.50327		0.496	0.48321	0.51907				0.52174	**5**
8	7475257	7479269						0.51241					**1**
8	9858995	10494049	0.48275		0.53167					0.4726	0.49561		**4**
8	10494122	11423794			0.5275	0.46581					0.50402		**3**
8	11433716	12250178			0.515	0.47488							**2**
8	13134093	13986575	0.47796		0.52667	0.46095							**3**
8	14047352	15341572	0.48633		0.51667								**2**
8	15341697	15344039	0.46522		0.52708	0.46639							**3**
8	15672318	16877193	0.46543		0.51208	0.48057							**3**
8	19731626	20188425			0.48667			0.475					**2**
8	20212901	20454380	0.4854		0.48125							0.46166	**3**

Chr	Start Chr Position	End Chr Position	MDI-landraces	MDI-impved varieties	Vietnam	Cambodia	Thailand	Philippines	Lao	Indonesia	Malaysia	Myanmar	**Groups of population**
8	24479317	24756546	0.49324	0.51634	0.515	0.46391	0.49997		0.49121		0.486	0.46756	**8**
8	24756591	25731670			0.45333								**1**
8	26214385	26407352	0.47247			0.46593	0.45478						**3**
8	27336079	27444789		0.47908									**1**
9	4208001	7432908		0.47582									**1**
9	8760534	9642993			0.45778					0.4609			**2**
9	11071329	11796090							0.47259			0.46462	**2**
9	11796099	11977311		0.46144									**1**
9	12120604	12120641	0.48654		0.46667	0.49697		0.47935					**4**
9	15158732	15508881			0.52583	0.507							**2**
9	20829666	21496031							0.4654				**1**
10	3795830	4017585									0.45096		**1**
10	4222311	4655647		0.4549									**1**
10	4660006	5179791										0.45257	**1**
10	5847780	6136790			0.51167	0.47912							**2**
10	6160262	6303176		0.50654		0.45127			0.4889			0.45257	**4**
10	9054963	9914992				0.45903							**1**
10	9914993	10733713		0.45817				0.49447				0.47866	**3**
10	10959003	11196737		0.46601				0.51775		0.48353		0.46275	**4**
10	12204147	12713101		0.48497	0.48333		0.47368	0.45124	0.45932	0.49019			**6**
10	12979745	13000779		0.47516	0.51583					0.49113			**3**
10	13866255	14798353						0.46232					**1**
10	22322998	22955577	0.45997										**1**
11	2403719	2637143									0.47624		**1**
11	3550757	3550814			0.525	0.50529					0.47718		**3**
11	3550873	4426617			0.51042	0.4647							**2**
11	7115419	7115521			0.48								**1**
11	8450433	9066992									0.47342		**1**
11	14368106	14998373			0.455								**1**

Chr	Start Chr Position	End Chr Position	MDI-landraces	MDI-impved varieties	Vietnam	Cambodia	Thailand	Philippines	Lao	Indonesia	Malaysia	Myanmar	**Groups of population**
11	16733767	17020000			0.45167								**1**
11	17466912	18312801		0.46209			0.47835					0.51871	**3**
11	18984753	20727445		0.45294									**1**
11	22246071	23359488				0.47587	0.47231						**2**
11	23359505	23359590					0.47054			0.45921	0.48438	0.51002	**4**
11	23359624	23901205					0.47459	0.49656		0.45368	0.47052	0.48755	**5**
11	26265728	26437153	0.48713		0.51583	0.46333				0.47255	0.48949		**5**
11	27158523	27207478				0.47572	0.4556						**2**
12	3622824	3782181		0.50261									**1**
12	3827769	3827885		0.48693									**1**
12	7666829	8036070					0.49383		0.46119			0.49229	**3**
12	9246567	10110729					0.49438	0.51187		0.47247	0.49469		**4**
12	14148662	14768070	0.45281	0.48644	0.48333	0.50256	0.48596	0.46063	0.50786	0.49271	0.45398	0.46126	**10**
12	14768079	15181603	0.46426	0.48105	0.5125				0.46633	0.492	0.50495		**6**
12	15757668	15834808					0.4843	0.45833	0.49561				**3**
12	16504611	16504667							0.50814				**1**
12	16504720	16504850							0.48736				**1**
12	16504854	16507787							0.50514				**1**
12	17952534	18275896							0.45319	0.45422			**2**
12	18275911	18368843							0.49739	0.46378			**2**
12	23516570	24093499							0.47331				**1**
	Total		**30**	**52**	**46**	**38**	**35**	**47**	**30**	**32**	**34**	**43**	

Note: High nucleotide diversity (π-value > 0.45); totally 493 accessions consist of 81 MDI landraces, 412 rice accessions from 8 countries; specific high: only high diversity in MDI landrace accessions; common high: there were at least 7 out of 9 country groups, which were high diversity.

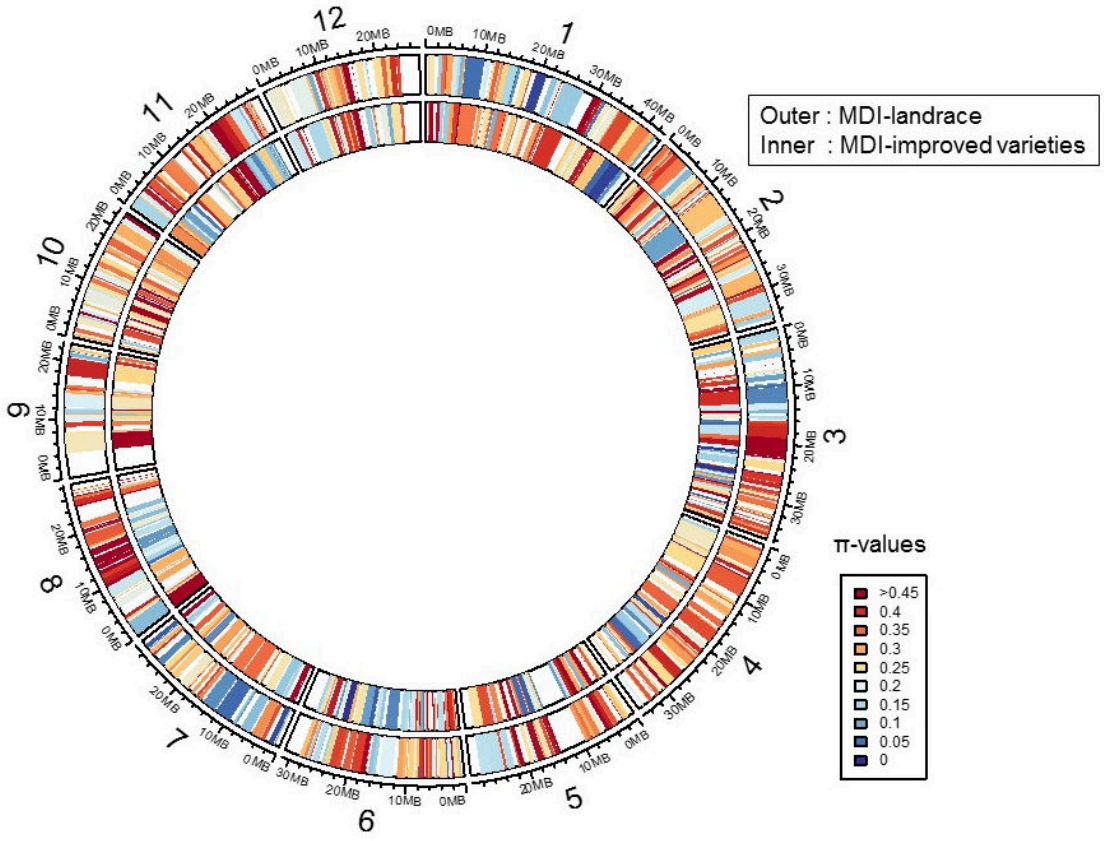
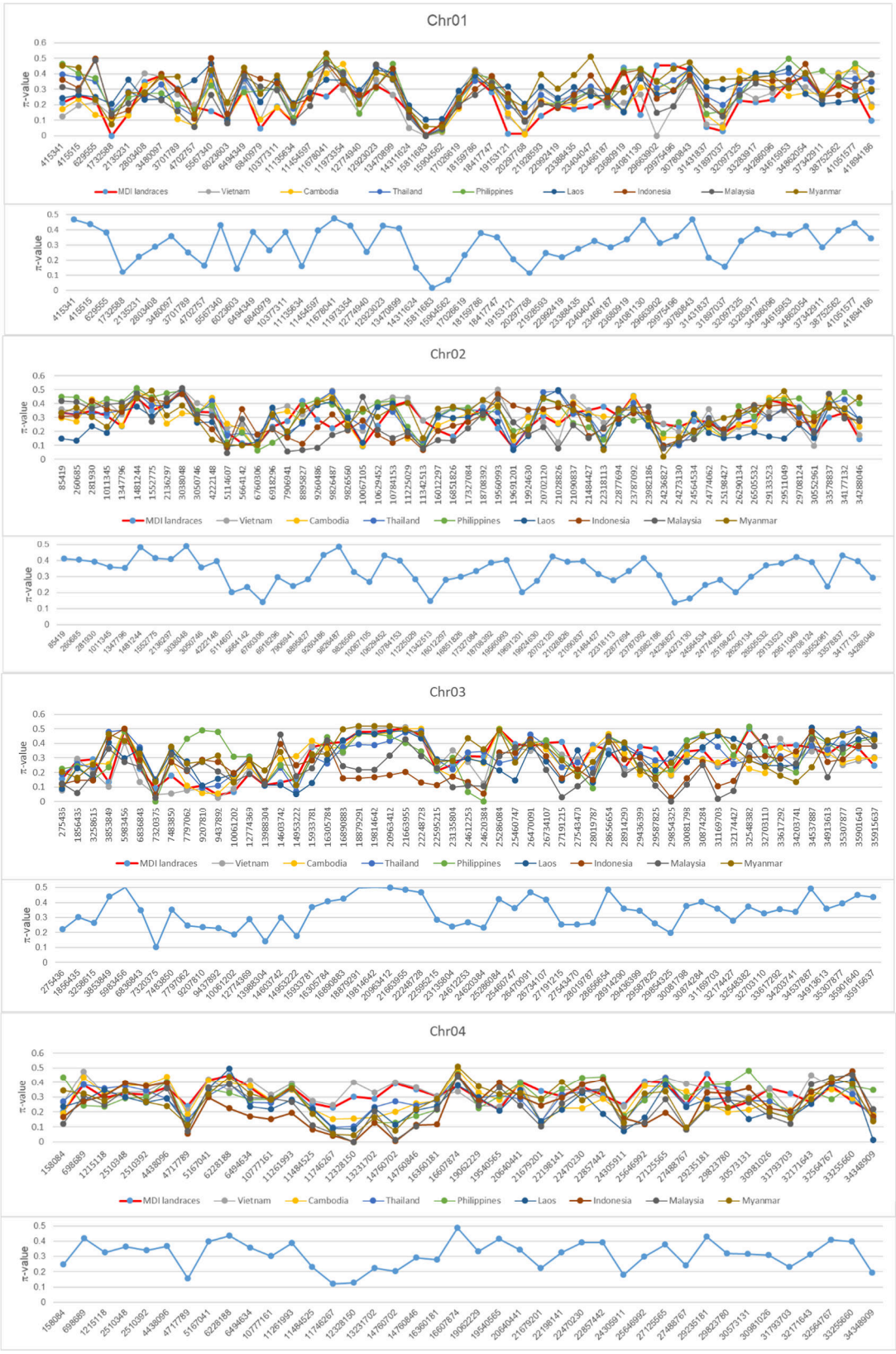
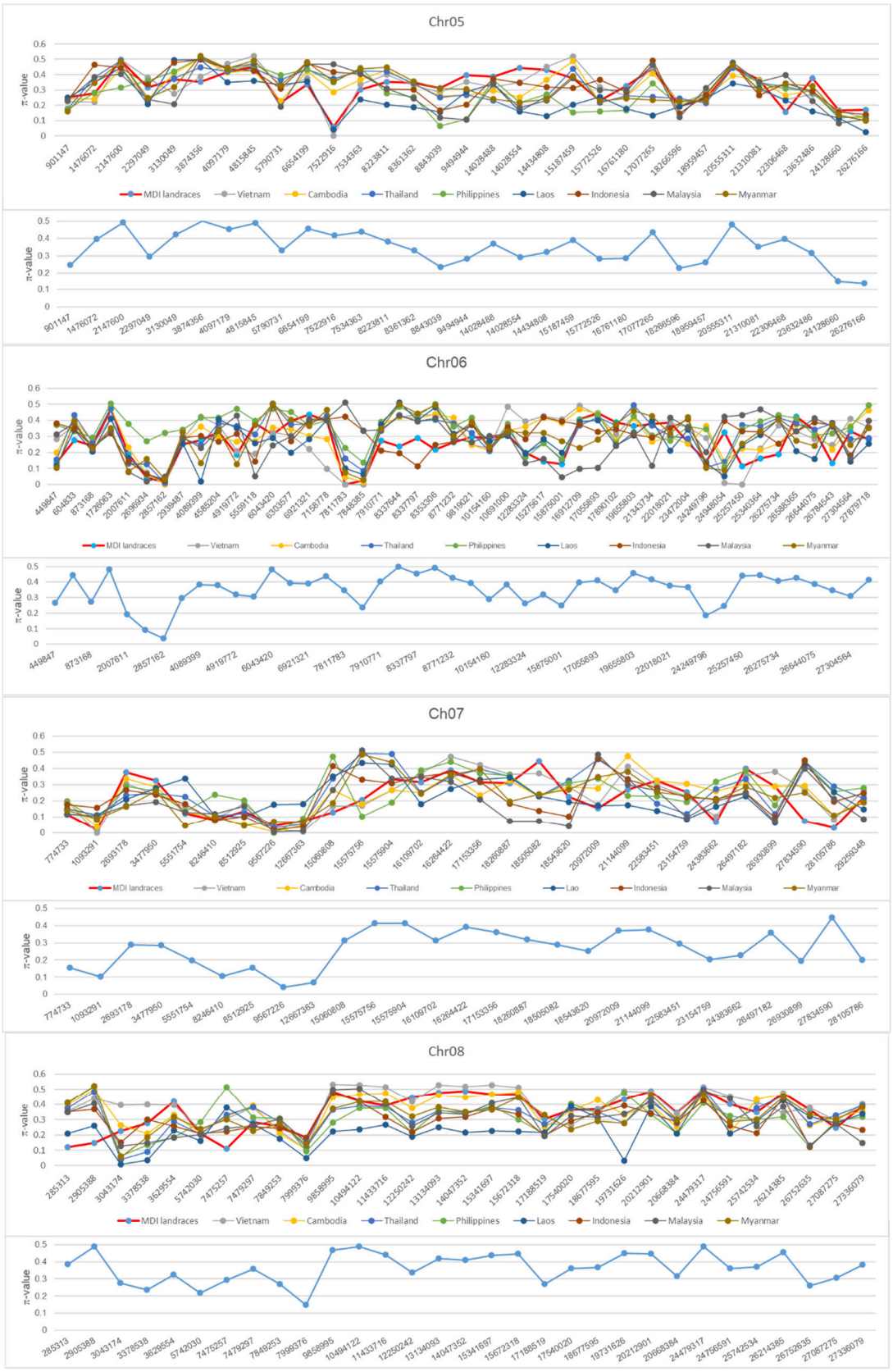


Figure 2.7. Circular genomic profiles of nucleotide diversities of MDI rice genomes.

Outer and inner circles correspond to MDI landraces and MDI improved varieties, respectively. Numbers 1 to 12 refer to the 12 chromosomes of the rice genome, with each tick mark corresponding to a 2-Mb interval. Nucleotide diversity levels (π-values) are color-coded as follows: red, high diversity; blue, low diversity.


Chr01
π-value
MDI landraces
Vietnam
Cambodia
Thailand
Philippines
Laos
Indonesia
Malaysia
Myanmar
Chr02
Chr03
Chr04


Chr05
π-value
MDI landraces
Vietnam
Cambodia
Thailand
Philippines
Laos
Indonesia
Malaysia
Myanmar
Chr06
Ch07
Chr08

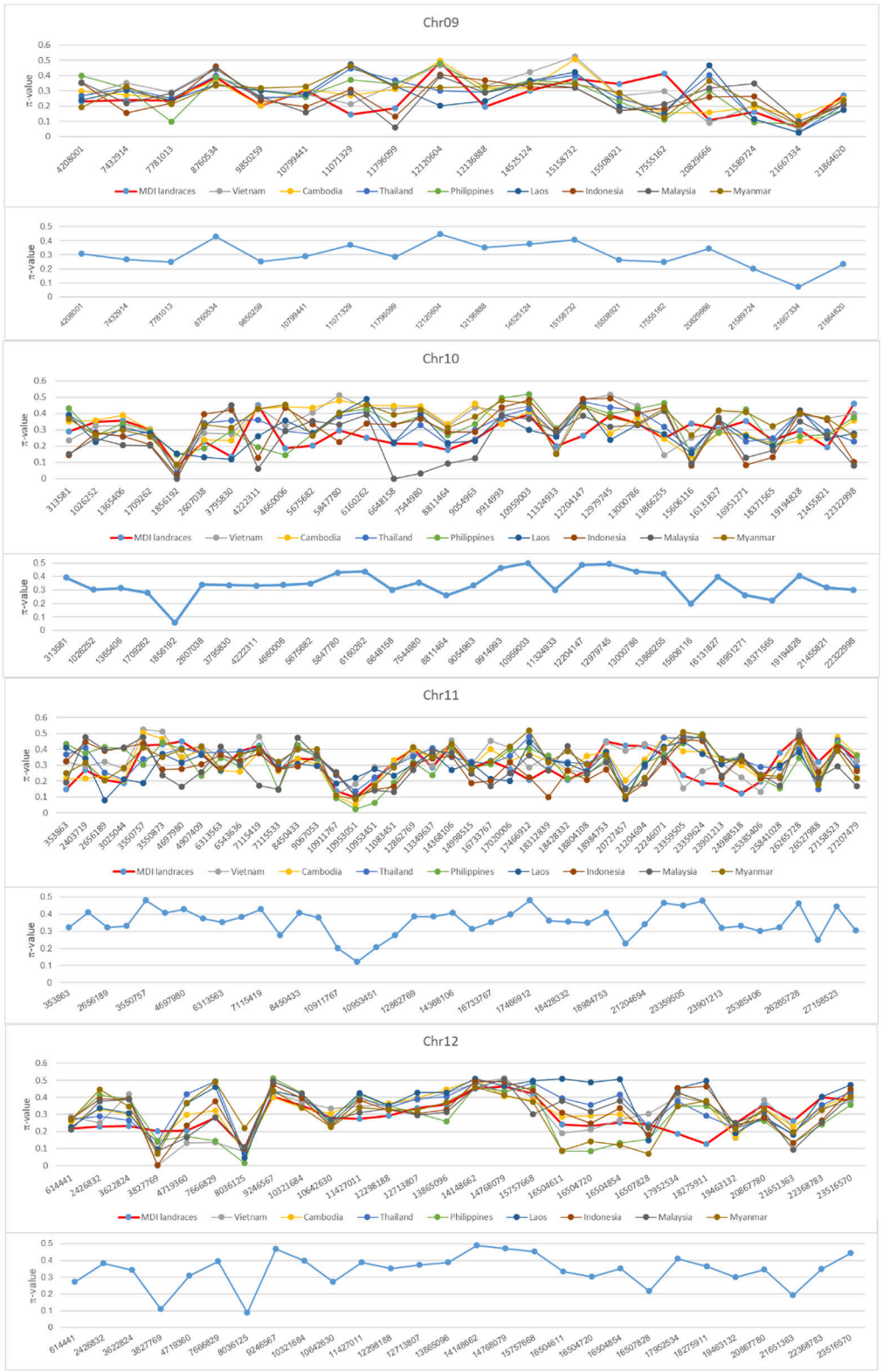


Figure 2.8. The π-value distribution of eight subpopulations and MDI landraces through overall 459 chromosomal regions.

The distribution is separately displayed in each of the 12 chromosomes. All π-values were calculated by the sliding window method in TASSEL v5.2.43 using the diversity tool with step and window sizes of five SNPs across 2,301 SNPs. In panels for each chromosome, the upper panel displays the π-value distributions for the nine subpopulations, and the lower panel shows the π-value distributions for a whole of 493 accessions from the nine subpopulations. X and Y axes indicate the nucleotide positions of SNPs in the chromosomes and π-values, respectively.

Table 2.5 Pairwise correlation of genome-wide nucleotide diversity of ind3 accessions based on country

	MDI-landraces	Vietnam	Cambodia	Thailand	Myanmar	Philippines	Laos	Malaysia
Vietnam	0.69							
Cambodia	0.64	0.77						
Thailand	0.46	0.51	0.68					
Myanmar	0.41	0.47	0.61	0.77				
Philippines	0.34	0.43	0.53	0.63	0.58			
Laos	0.3	0.35	0.5	0.76	0.62	0.46		
Malaysia	0.26	0.33	0.43	0.59	0.48	0.52	0.37	
Indonesia	0.25	0.36	0.47	0.58	0.51	0.55	0.38	0.76

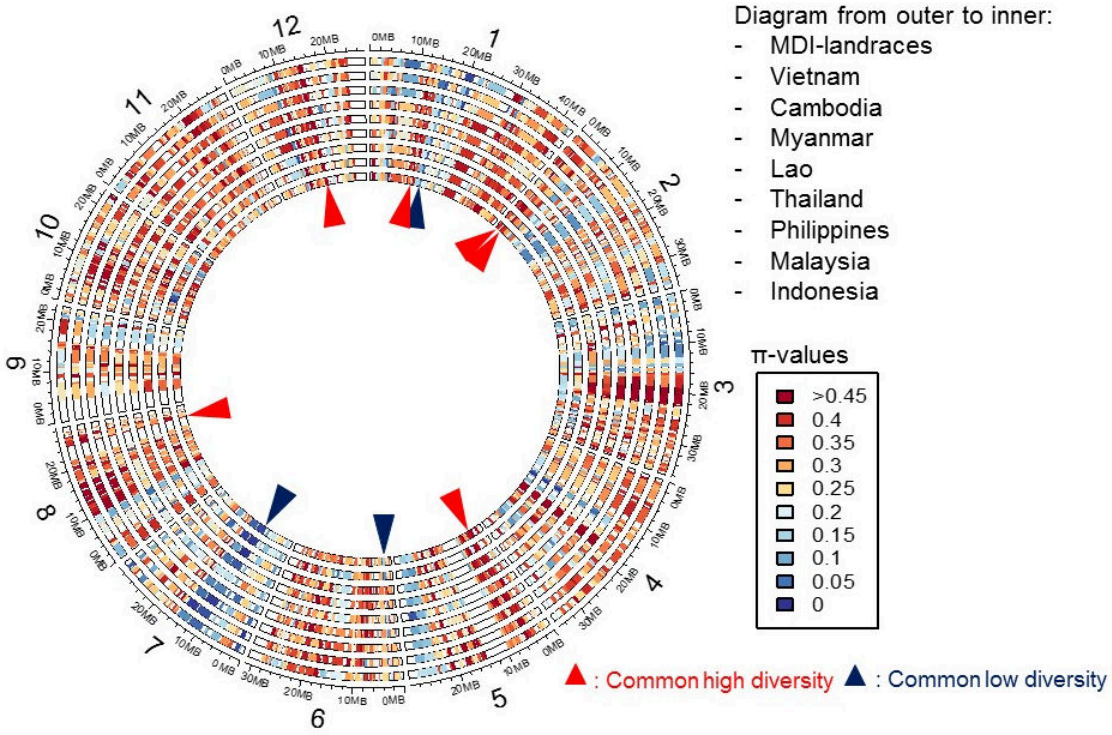


Figure 2.9. Circular profiles of nucleotide diversities of MDI landraces genomes and Ind3 accessions from Southeast Asian countries

Numbers 1 to 12 refer to the 12 chromosomes of the rice genome, with each tick mark corresponding to a 2-Mb interval. The color scale, which ranges from deep blue (<0.05) to red (>0.45), indicates the level of nucleotide diversity based on π-values. Red and blue arrows respectively indicate high- and low-diversity regions shared by six of eight groups representing different Southeast Asian countries.

Table 2.6. Genes known for common low diversity of ind3 from eight Southeast Asian countries and MDI-landraces

Chr.	Name	Description (protein/locus name)	Position	Database
01 15811683-15904518	LOC_Os01g28230	ZOS1-07 - C_2H_2 zinc finger protein	15814357-15815151	RAP-DB
	LOC_Os01g28300	OsFBX7 - F-box domain containing protein	15850967-15854629	RAP-DB
06 2857162-2939468	Os06g0157500-RFT1	RICE FLOWERING LOCUS T 1	2926823-2928474	RAP-DB RICE SNP-SEEK DATABASE
07 9567226-12431967	No gene was identified	-	-	RICE SNP-SEEK DATABASE

Note: these genes were getting by searching trait genes in website Rice SNP-Seek Database (http://snp-seek.irri.org).

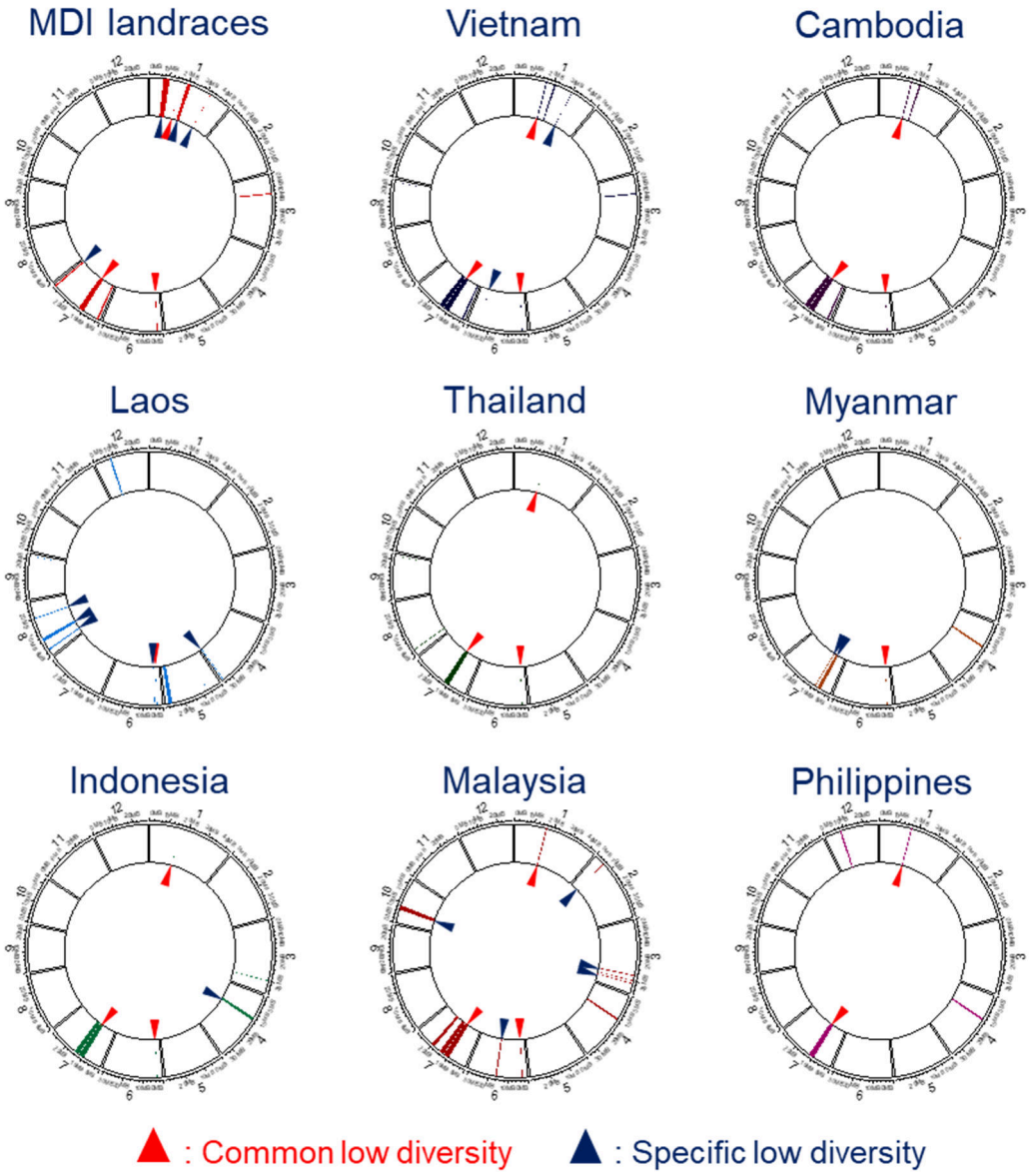


Figure 2.10. Simple circular profiles of low diversity in MDI landrace genomes and Ind3 accessions from Southeast Asian countries

Red and blue arrows point to regions of low diversity ($\pi<0.05$) that are respectively common to six of eight Southeast Asian groups or population specific.

Chapter 3. The genome-wide association study of salinity responses

3.1. Introduction

Rice (*Oryza sativa* L.) is the main staple foods source for about 90% of Asian people (Ghomi et al., 2013, Hariadi et al., 2015) and is grown in the areas where it can access water (Batayeva et al., 2018). One of important threats of rice production currently and in the future is salinity stress (Foti et al., 2019). Salinity stress can be from salt intrusion from the sea in the dry season. In coastal and/or acid and semi-arid areas, the effect of salt conditions on rice production is a major issue (Batayeva et al., 2018). Particularly, about 50,000 ha of rice land would be affected by salinity intrusion under the climate change and sea level rise scenario by 2050 and even 10 times more under the worst case of the combined climate change, sea level rise and low river flows of the Mekong Delta/River (Kontgis et al., 2019, Nhung et al., 2019). Under such cases, a huge yield loss will happen, an estimated range of 100,000 – 1 million tons of paddies per year (Smajgl *et al.,* 2014). Some measures are proposed to deal with salinity problem such as soft measures (i.e. salinity-tolerant rice varieties, adaptive rice farming practices, adaptive rice-based farming systems, replacing rice by another more adaptive crop), hard measure (salinity-manage structures), or the combined soft and hard measures. The use of salinity-tolerant varieties is considered as the most efficient and effective way (Emon et al., 2015).

Rice is relatively sensitive to salt condition (Krishnamurthy et al., 2011, Nakhoda et al., 2012, Ghomi et al., 2013, Platten et al., 2013), especially at seedling stage (Zheng et al., 2015, Batayeva et al., 2018). Therefore, a lot of previous studies focused on screening and selection tolerant rice varieties on this period (Lee et al., 2003, Kumari et al., 2008, Genc et al., 2010, Tian et al., 2011, De Leon et al., 2016, Puram et al., 2017, De Leon et al., 2017). There are several protocols developed for screening salinity tolerance at seedling stage. For example, growing rice varieties in nutrient solution containing 50 mM NaCl for four weeks before evaluation (Frouin et al., 2018), screening rice varieties in salt on concentration 6 dSm^{-1} in two days, and change to salt

concentration of 12 dSm^{-1} in four days (De Leon et al., 2017), or growing rice in soils having four different salt level (50 mM, 100 mM, 150 mM and 200 mM of NaCl (Hariadi et al., 2015).

Ahmadi et al. (2011) used 124 SNP and 52 SSR markers to identify 14 quantitative trait loci (QTLs) and 65 candidate genes related to the phenotype of salinity tolerance of 180 japonica accessions. Yu et al. (2017) used 295 accessions (218 *japonica* accessions, 72 *indica* accessions and five admixture accessions) to perform GWAS and they identified 93 candidate genes related to salinity tolerance at seedling stage. In addition, Shi et al. (2017) studied on salinity tolerance at the seed germination stage and applied GWAS for understanding the correlation of SNP candidates and the accessions that they could service in salt condition.

Rice plant growth responds to salinity stress via two phases: osmotic phase and ionic phase (Munns & Tester, 2008). In salt stress, plant is rapidly affected by osmotic pressure that inhibits growth of young leaves. If plant grows longer in salt condition, it will face the ionic problematic issues that accelerates senescence of mature leaves (Munns & Tester, 2008). Therefore, rice plant can adapt to salt stress in three distinct ways: osmotic stress tolerance, Na^+ and/or Cl^- exclusion, and tolerance on tissue level with accumulated Na^+ and/or Cl^- (Munns & Tester, 2008).

Currently, about 2,000 samples of local and improved rice varieties in the Mekong Delta have been collected and conserved at the Mekong Delta Development Research Institute of Can Tho University (Vietnam) as the genetic material for salinity-tolerant rice breeding. Among these, 99 rice accessions were selected for salinity tolerance screening, and to identify genetic basis of salinity tolerance in MDI rice. Therefore, the evaluation and classification of rice varieties, which can tolerate salty soil conditions, is one of the solutions to deal with climate change in the future. This study aimed to evaluate a set of MDI rice accessions related to salinity tolerance at seedling stage. Moreover, I would like to identify the relationship between genotype and phenotype on salinity tolerance related traits by using SNPs from ddRAD-seq.

3.2. Materials and methods

3.2.1. Plant materials

Ninety-nine rice accessions consisted of 81 landraces and 18 improved varieties chosen from the genebank of Mekong Delta Development Research Institute – Can Tho University, Vietnam (Table 2.1).

3.2.2. Screening for salinity tolerance

Seeds were put in 20 percent of Tekurindo C Flowable solution (Kumiai Kagaku Kogyo company) at 15 minutes for sterilization and then washed with tap water (3 times), seeds were then washed with distilled water 2 times to remove chemical. Sterilized seeds were put in petri dish and germinated in incubator. Three days after germination, sow those seeds in net sheet floating on distilled water in a 40-L container in green house. 4 days after sowing (DAS), solution was replaced by the standard nutrient solution. The standard nutrient solution contained 1.12 mM NH_4NO_3, 0.32 mM $NaH_2PO_4.2H_2O$, 0.19 mM K_2SO_4, 0.38 mM KCl, 1.25 mM $CaCl_2.2H_2O$, 0.82 mM $MgSO_4.7H_2O$, 35.8 µM Fe-EDTA, 9.1 µM $Mn_2SO_4.5H_2O$, 3.06 µM $ZnSO_4.7H_2O$, 0.16 µM $CuSO_4.5H_2O$, 46.2 µM H_3BO_3, 0.052 µM $(NH_4)_6Mo_7O_{24}.4H_2O$. The solution was renewed every three days and pH was adjusted to 5.0 every day by 1M NaOH and 0.5M HCl. At 15 DAS, the salt stress treatment was started (add 100 mM NaCl to the nutrient solution). Just before treatment, measure the plant height and root length of all samples. 5 days after treatment (DAT), I measured plant height, root length and evaluated the visual reactions of plant to salinity stress using modified Standard Evaluation System (SES) Score by IRRI (2014) to access the salinity tolerance of each variety (Table 3.1). At 20 DAT, I scored the SES, measured plant height, root length, and harvested samples. All plants were washed with tap water (2 times) and then distilled water 2 times to remove salt on plant surface. Plants were separated into two parts (shoot and root), dried at 70^0C for 80 h and weighed. The process for salinity screening was showed in Fig. 3.1.

Table 3.1. Phenotypic features for the Standard Evaluation System (SES) scores

Score	Status	Phenotypic features
1	High tolerance	Growth and tillering nearly normal
2	High tolerance	Growth nearly normal but there is some reduction in tillering and one leaf whitish and rolled
3	Tolerance	Growth nearly normal but there is some reduction in tillering and two leaves whitish and rolled
4	Tolerance	Growth and tillering reduced; three-four leaves whitish and rolled (salt); only a few elongating
5	Moderately tolerance	Growth and tillering reduced; most leaves whitish and rolled (salt); only a few elongating
6	Moderately tolerance	Growth completely ceases; most leaves dry; the latest leaf whitish
7	Susceptible	Growth completely ceases; most leaves dry; plant dying
8	Susceptible	Plant dying totally
9	High susceptible	Plant dead and dry

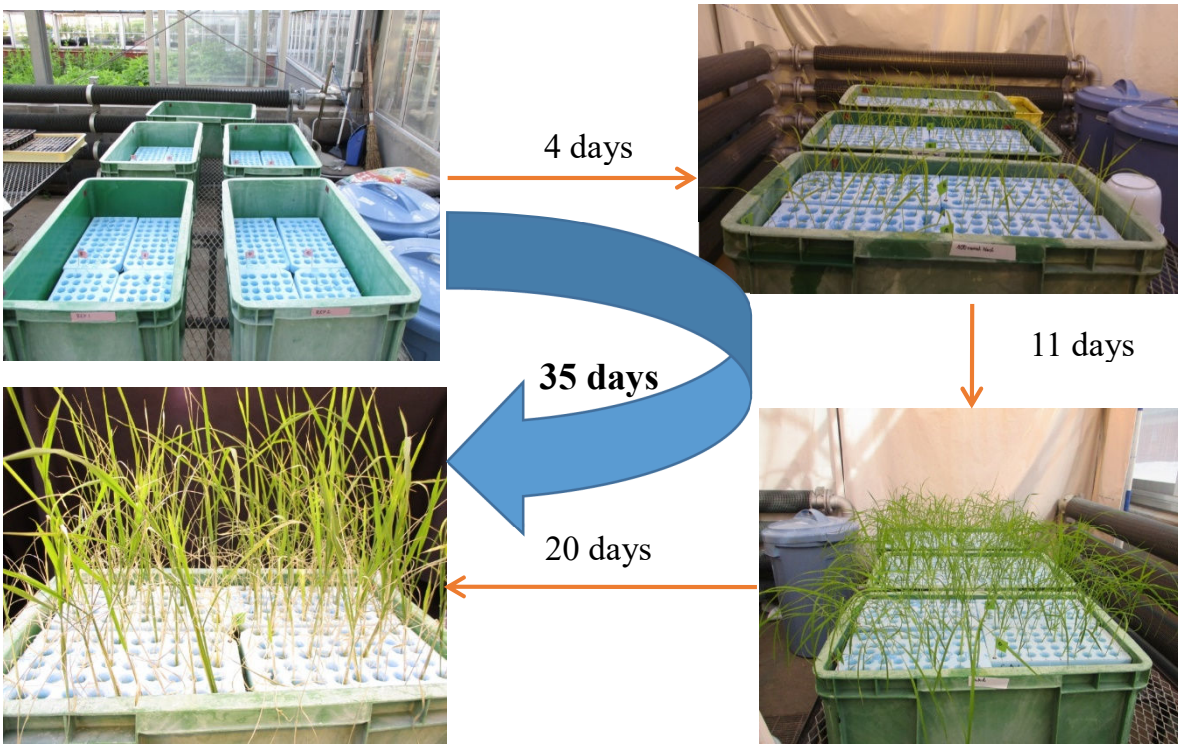


Figure 3.1. The process for salinity treatment on nutrient solution

(A) Germinated seeds were sown in net sheet floating on distilled water; (B) plants in net sheet floating put on nutrient solution; (C) plants in net sheet floating put on nutrient solution with 100 mM NaCl; (D) Plants after treated 20 days on nutrient solution with 100 mM NaCl.

3.2.3. Genome sequence data analysis

SNPs distributed across 12 chromosomes for 99 rice varieties in RAD-seq dataset (578,704 SNPs) were filtered based on minimum call rate 60%; heterozygous rate 0.2, minor allele frequency 0.02, resulted in 37,643 SNPs for imputation. The imputation was performed using Beagle v5.0 (Browning et al., 2018).

3.2.4. Data analysis

For the phenotypic data of salinity tolerance related traits, the mean value of three replications was calculated and used for genome-wide association study (GWAS).

The relative elongation of plant height and root length were calculated as follows:

$$R = \frac{A}{B}$$

R: The relative elongation

A: (Plant height/root length 20 DAT – Plant height/root length 0 DAT) of NaCl treatment

B: (Plant height/root length 20 DAT – Plant height/root length 0 DAT) of control treatment.

I used TASSEL 5.2.50 (Trait Analysis by Association Evolution and Linkage) to filter, perform principal components analysis (PCA), general Linear model (GLM). The significant threshold was set at $p < 0.0001$ ($-\log_{10} p$-value >4.0). Significant level of marker-trait association ($-\log_{10} p$-value) was also assessed using q-value (FDE adjusted p-value) by using R software (Dabney et al., 2004). SNPs with the q-value lower than 0.05 was selected as significant marker. Candidate genes (near peak SNPs) were searched on QTL and gene database in the rice SNP seek database (Mansueto et al., 2016).

3.3. Results

3.3.1. Salinity tolerance of selected rice accessions

The salinity tolerance among 99 accessions was evaluated at 5, 10, 15 and 20 days after treatment (DAT). Two accessions (MDI-93 and MDI-124) did not successfully germinate. Therefore, I tested 97 rice accessions. The detailed information of SES score and four phenotypic traits in each accession can be seen in Supplementary Table 1. Variation of salt tolerance among different varieties was large. Even at 20 DAT, five accessions showed the SES score of ≤3, and 47 accessions had the SES of 3.0-5.0. By contrast 28 accessions were the SES score of 8-9. As the treatment period increased, the SES score tended to increase. At 5 DAT, the SES score of ≤3.0 was about 66%, and about 34% of accessions was for SES score of 3.0-5.0. At 10 DAT, the proportion of these accessions changed with about 28.9% (28 accessions), 57.7% (56 accessions), 11.3% (11 accessions) which was high tolerance, tolerance, and moderate, respectively. In particular, two accessions reached a score level at 7.0-8.0 (sensitive). At 15 DAT, the number of sensitive varieties were 10 accessions, which accounted for 10.3%, whereas, the accessions at tolerant group was 14.4% (14 accessions) (Fig. 3.2).

3.3.2. Phenotypic characterization in salt condition

Plant height and root length elongation

There were significant differences in four phenotypic traits and SES at 20 DAT (Fig. 3.3). Group with small SES score showed the higher relative elongation on plant height and root length or the shoot dry weight and root dry weight. The shoot elongation was the most affected by the NaCl condition, whereas root length elongation ratio was the least affected (Fig. 3.3). Regarding the biomass of rice accessions at seedling stage between two treatments (with 100 mM NaCl and without NaCl) and seven SES categories. Shoot and root dry weight were the same trend in seven SES score groups, the highest ratio between NaCl treatment and control treatment was in SES score≤3.0, which were about 0.5, but there were more variety on root dry weight than shoot dry weight. Following were in SES score 3.0-4.0 in both shoot and root dry weight. The

lowest one was in SES score 8.0-9.0, which were about 0.2 for shoot dry and root dry weight (Fig. 3.3).

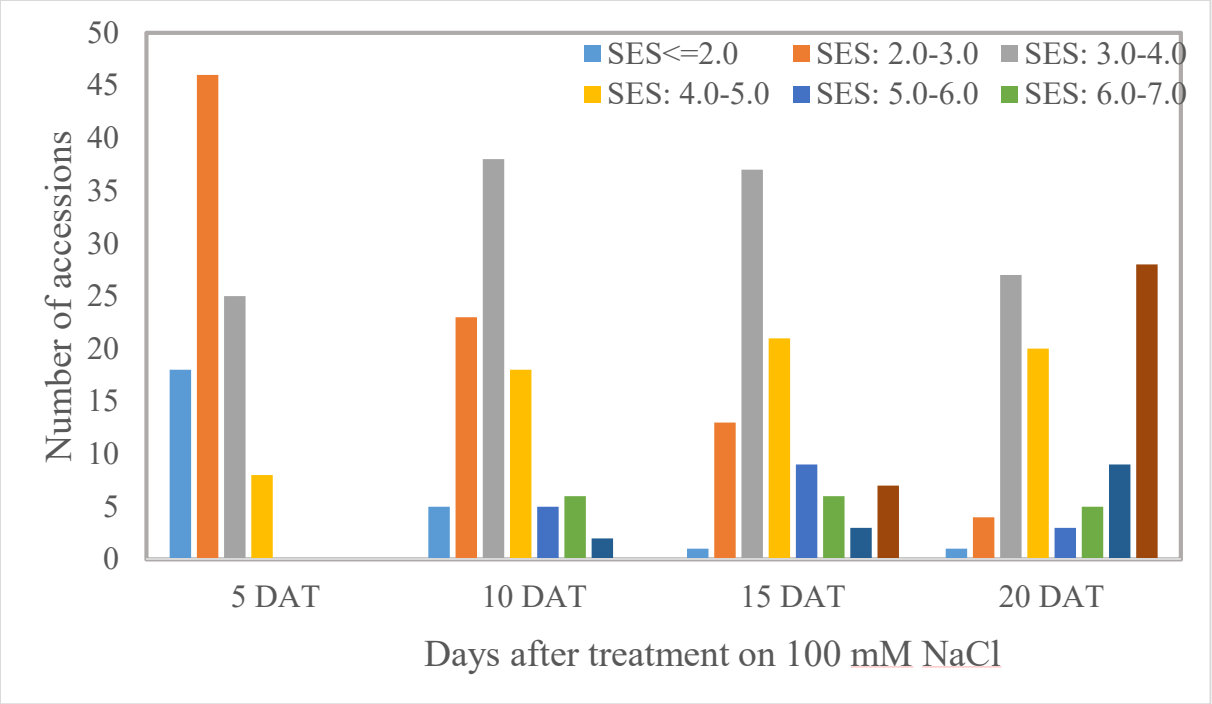


Fig 3.2. Number of accessions in difference categories

Total 97 rice accessions with 3 replications were treated in nutrient solution which added 100 mM NaCl at 15 days after sowing. The salinity tolerance was evaluated and recorded at 5, 10, 15, and 20 days after treated with sodium chloride (DAT), respectively. SES score value was divided into eight categories: 1: ≤2.0; 2: 2.0-3.0; 3: 3.0-4.0; 4: 4.0-5.0; 5: 5.0-6.0; 6: 6.0-7.0; 7: 7.0-8.0; 8: 8.0-9.0. Five accessions had high tolerant with salt condition were Doc Phung, Trang Phieu, Nang Quot Bien, Mot Bui Lun and Lun Can Trang, respectively.

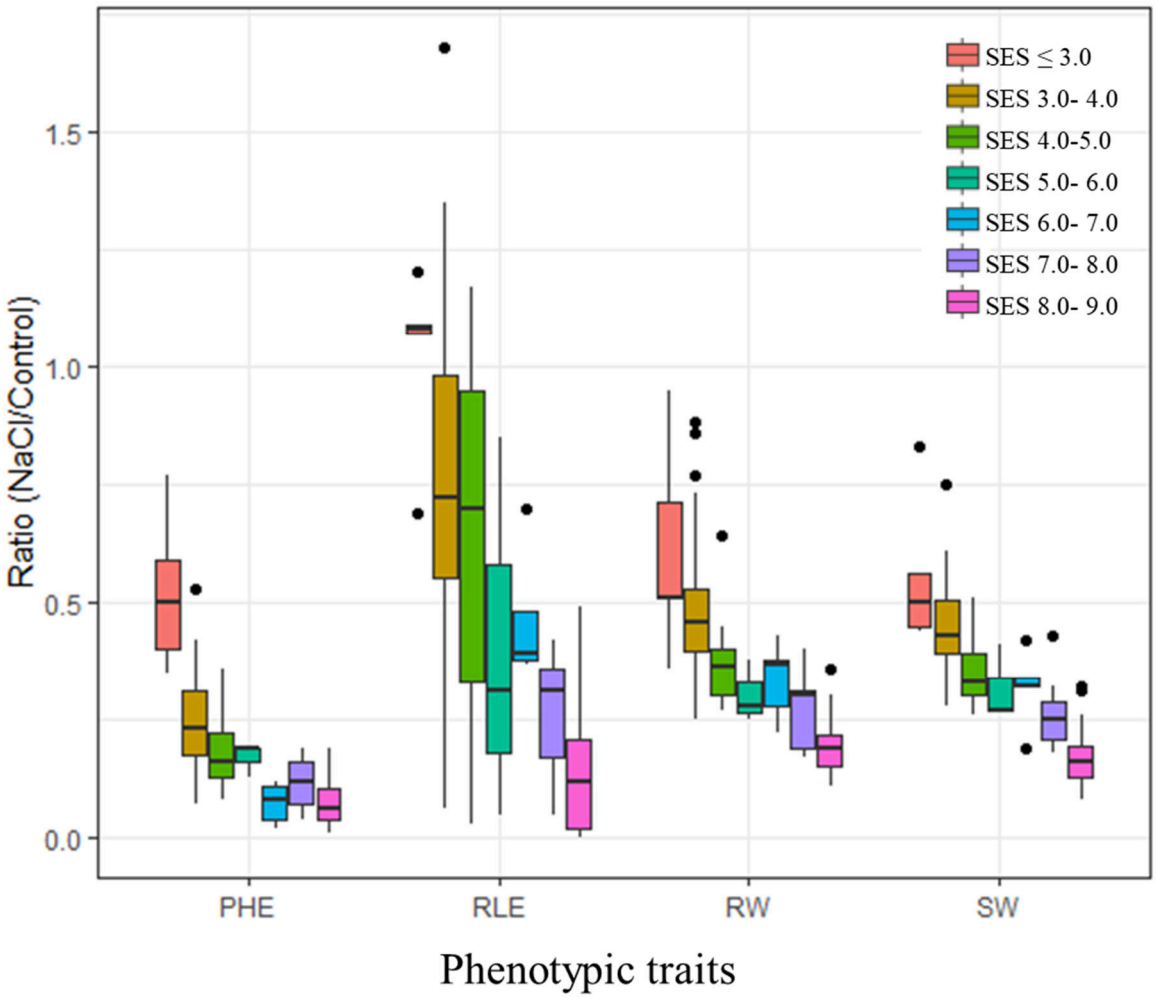


Fig 3.3. Effect of salt stress on plant morphology traits

PHE: plant height elongation; RLE: root length elongation; RW: root dry weight; SW: shoot dry weight.

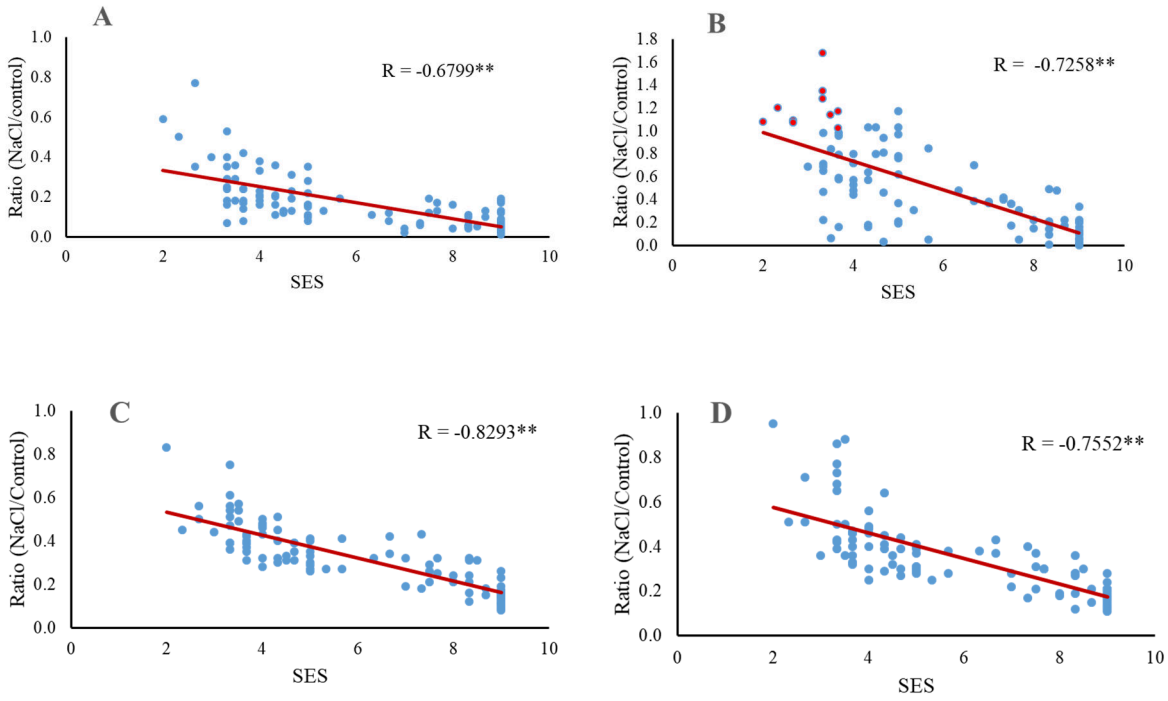


Figure 3.4. Correlation between SES score 20 DAT and the ratio of other salinity traits between NaCl treatment and control treatment

*A) Plant height elongation ratio; B) Root length elongation ratio; C) Shoot dry weight ratio; D) Root dry weight ratio; **. Correlation is significant at the 0.01 level.*

Correlation of phenotypic traits of 97 accessions

The SES score was significantly and negatively correlated with all the evaluated traits at 20 DAT (Fig. 3.4). Among them, the Pearson correlation coefficient was highest on relative shoot dry weight (0.8293), followed by relative root dry weight (0.7552), the Pearson correlation coefficient was lowest on plant height elongation ratio (0.6799). In the relative root length elongation, 12 accessions had the value more than 1.0 (Fig. 3.4).

3.3.3. Genome wide association study (GWAS) between phenotype of traits related to salt tolerant rice accessions and genotype

GWAS for SES score among 99 rice accessions

Using 37,643 imputed SNPs based on ddRAD-seq data of 99 accessions and 20 whole genome accessions for reference and salinity traits of 99 rice accessions at four stages of treatment to perform GWAS, three genomic regions were significantly correlated with SES score at 20 DAT, which were located in Chr01, Chr04 and Chr10, respectively (Fig. 3.5).

Focusing on the threshold of $-\log_{10}$(p-value) which was more than 4.0, I identified 15 SNP markers. Five, three and seven SNP markers were located on Chr01, Chr04 and Chr10, respectively (Fig. 3.5, Table 3.2). Significant SNP threshold is defined as follows: (1) –log (p-value) > 4.0, (2) False discovery rate (q-value) < 0.05. Q-value was calculated using R-software. Only one SNP marker on chromosome 1 (S01_18015212) was significant at the q-value ≤0.05 (Table 3.2).

In addition, I also classified these SNPs marker which were appeared on other three conservational stages of screening (Fig. 3.6). In particularly, there were many significant SNPs which appeared the same location in chromosome 10 on four SES conservational stages (Fig. 3.6). Additionally, the SNP peak marker on chromosome 1 (S01_18015212) was appeared on three out of four treatment periods (5 DAT, 10 DAT, 20 DAT) (Fig. 3.6A, 3.6B, 3.6D). Meanwhile, it was the second SNP peak for treatment 15 DAT (Fig. 3.6C).

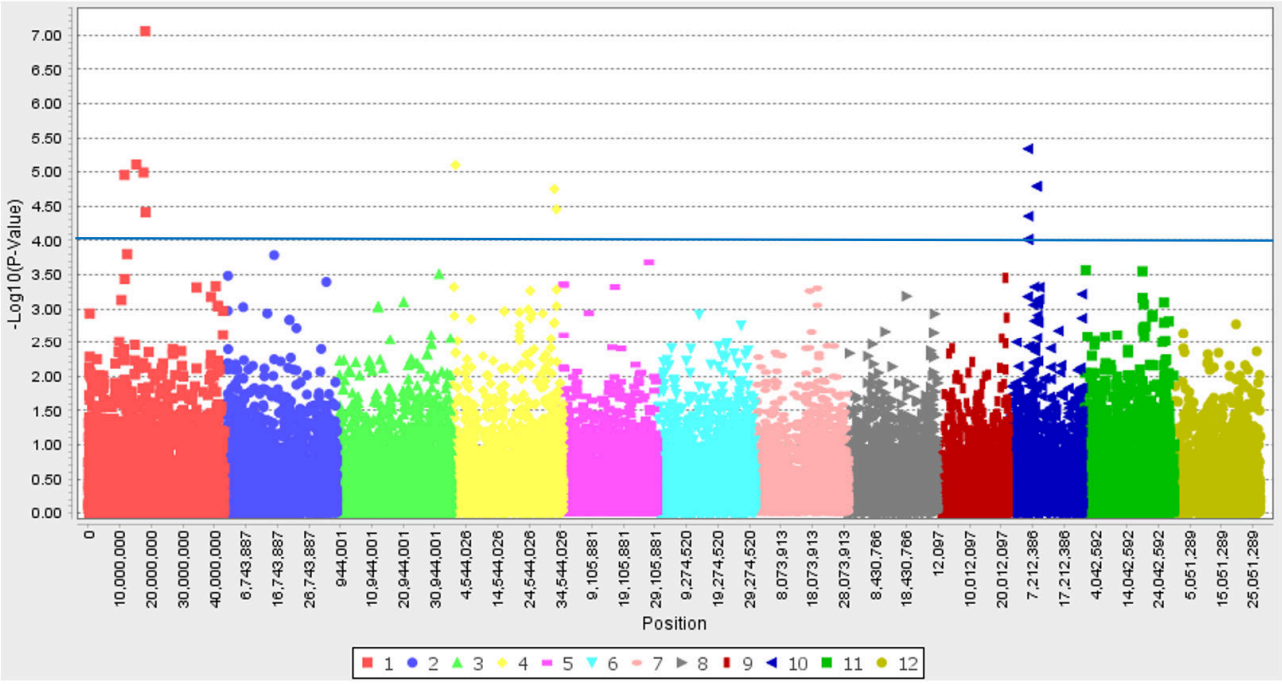


Fig 3.5. GWAS for salinity tolerance in 100 mM NaCl at 20 DAT

Manhattan plot for SES 20 DAT. Using 37,643 SNPs from imputation of 99 accessions and their SES score to perform GWAS. Three chromosomes were significant correlated with SES 20 DAT, namely chr01, chr04 and chr10. SNP marker at position 18,015,212bp in chromosome 01 was at the peak.

Table 3.2. Significant SNPs for SES 20 DAT

Marker	Chr	Position	p-value	q-value
S01_11466366	1	11466366	1.10E-05	0.06901
S01_15190150	1	15190150	7.74E-06	0.06901
S01_17479548	1	17479548	1.01E-05	0.06901
S01_18015212	1	18015212	8.77E-08	0.003301
S01_18151907	1	18151907	3.85E-05	0.131747
S04_1215043	4	1215043	7.89E-06	0.06901
S04_32564790	4	32564790	1.77E-05	0.074029
S04_33255725	4	33255725	3.50E-05	0.131747
S10_5805380	10	5805380	4.58E-06	0.06901
S10_5847722	10	5847722	9.67E-05	0.242665
S10_5847733	10	5847733	9.67E-05	0.242665
S10_5847780	10	5847780	9.67E-05	0.242665
S10_5928011	10	5928011	4.41E-05	0.138334
S10_8601860	10	8601860	1.61E-05	0.074029
S10_8601916	10	8601916	1.61E-05	0.074029

Significant SNPs at the level of q-value ≤0.05

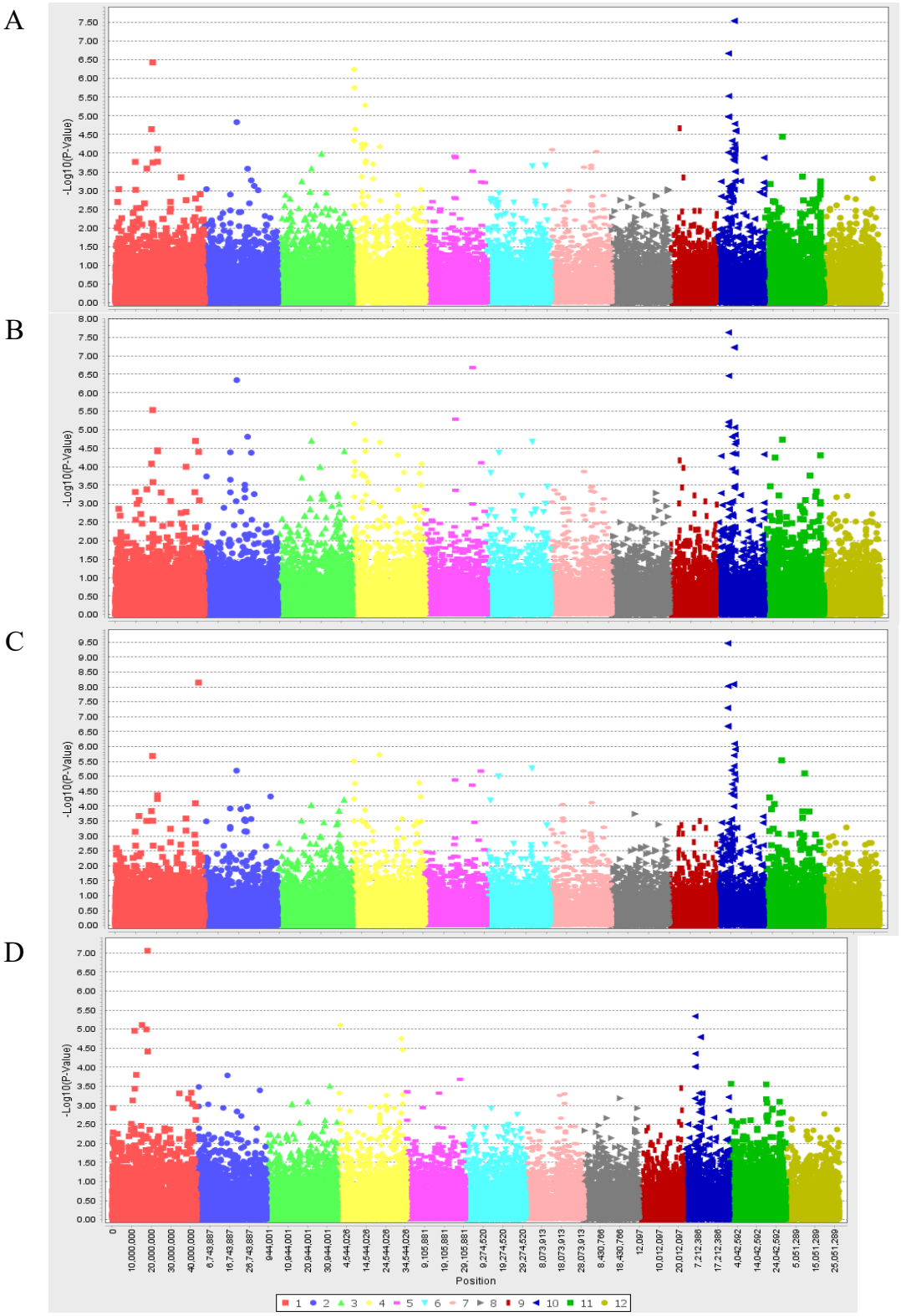


Figure 3.6. GWAS for salinity tolerance of plant in 100 mM NaCl at four stages of screening

(A) Manhattan plot for SES 5 DAT, (B) Manhattan plot for SES 10 DAT, (C) Manhattan plot for SES 15 DAT and (D) Manhattan plot for SES 20 DAT. Using 37,643 SNPs from 99 accessions and their SES score to perform GWAS. Three chromosomes were significant correlated with SES score at four stages of treatment, namely chr01, chr04 and chr10. SNP marker at position 18,015,212bp in chr01 was at the peak in three out of four stages. There were many significant SNPs in chr10 at four treated stages.

According the peak of significant SNPs in chromosome 01 (S01_18015212) (Fig. 3.7A), I filtered all SNPs information from 99 rice accessions with the upstream and downstream of the peak 100kb, total filtered 27 SNPs were used for Haploview analysis. The linkage disequilibrium (LD) block was out of the peak (Fig. 3.7 B), but the peak of SNP was in the 3' UTR of gene LOC_Os01g32830 (Fig. 3.7C). At the position of 18,015,212 bp, I identified 3 haplotypes which were C allele (major allele), A allele (minor allele) and M (heterozygous) (Fig. 3.8A, Table 3.3). The differences among three haplotypes were shown in Fig. 3.8A and Table 3.3, the SES score was low for accessions had the minor allele (3.9) or heterozygous (3.1), whereas, it was high for rice accessions had major allele (6.4) (Fig. 3.8A, Table 3.3).

In the length of 3,353bp of gene LOC_Os01g32830, I found 9 SNPs that were located in this gene from the dataset of 99 MDI rice accessions, but those of SNPs markers were not significant with SES 20 DAT from GWAS results. Only SNP marker S01_18015212 was high significant (Supplementary Table 2). The p-value also explained why the LD of their SNPs were not in the same haploblock. I also found the peak SNP (S01_18015212) and the peak S01_18015102 had high LD (Fig. 3.7B).

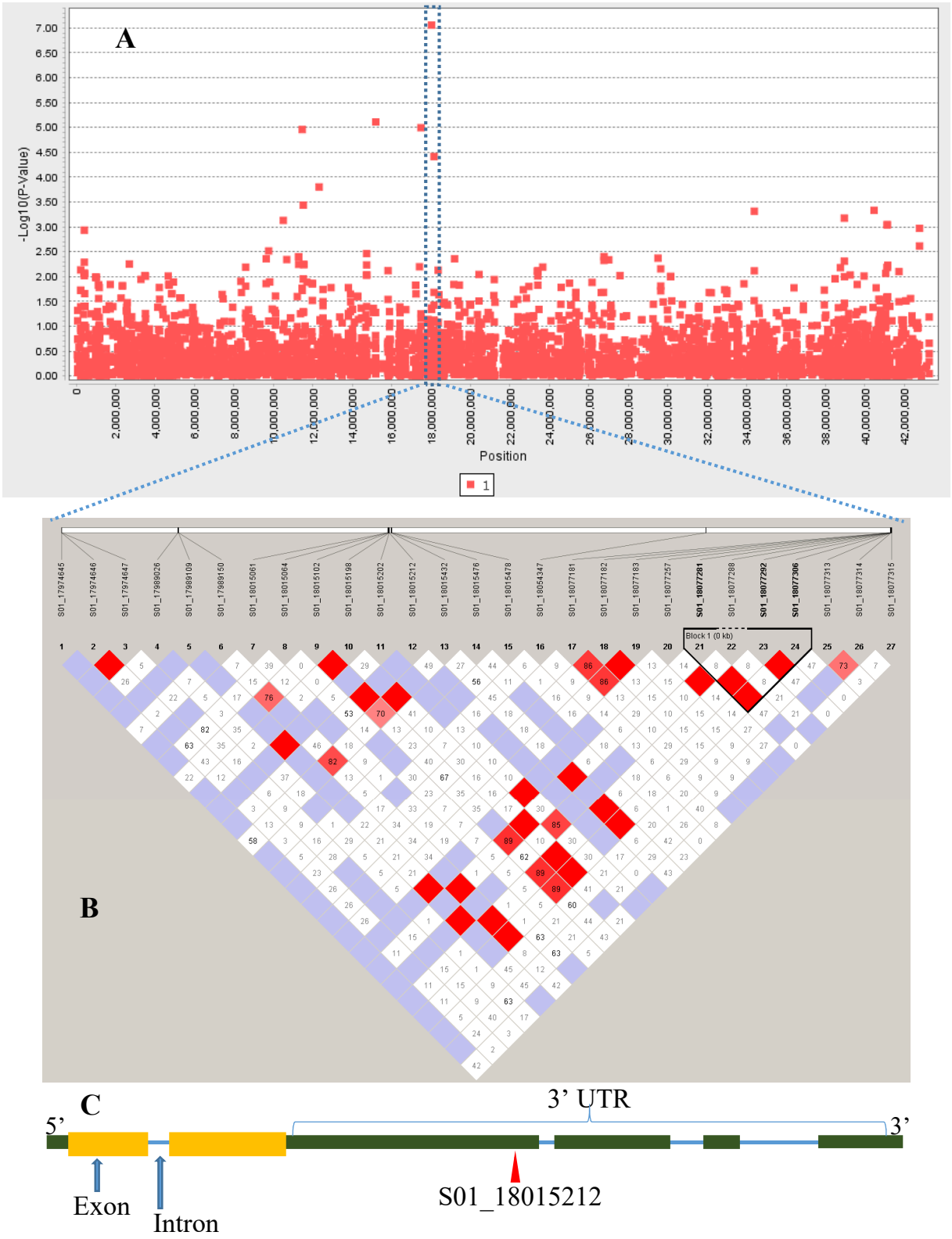


Figure 3.7. Analyses of the peak for salinity traits on chromosome 01

(A) Manhattan plot for SES score, (B) LD block of SNP peak and other, and (C) Gene association with SNP peak. SNP marker in the peak of chromosome 01 was located in 18015212 bp (S01_18015212)

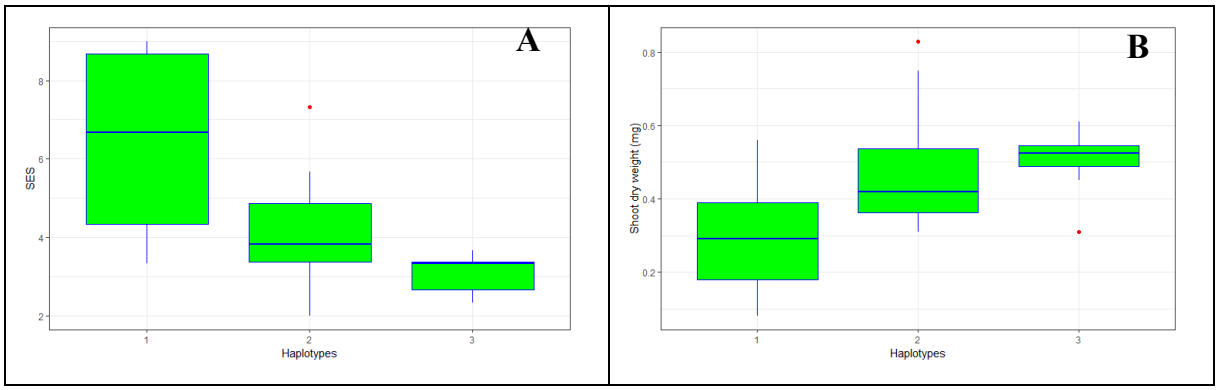


Figure 3.8. SES score and shoot dry weight in three distinct haplotypes on chromosome 01

(A) SES score for three haplotypes, (B) Shoot dry weight ratio for three haplotypes. SNP marker in the peak of chromosome 01 was located in 18015212 bp (S01_18015212)

Table 3.3. Haplotype information of SNP marker S01_18015212

Haplotype	Allele	Number of accessions	SES score	Shoot dry weight ratio
Hap1	C	80	6.4 ± 0.24^{a}	0.29 ± 0.01^{b}
Hap2	A	10	4.2 ± 0.48^{b}	0.48 ± 0.06^{a}
Hap3	M	9	3.1 ± 0.17^{b}	0.50 ± 0.03^{a}
Total		99		
P-value			$1.8e^{-6}$	$7.11e^{-8}$

SES score and shoot dry weight ratio were measured for 97 accessions. One accession in Hap1 and one accession in Hap3 were not germinate. Hap: Haplotype

Regarding the peak SNP marker in chromosome 04 (S04_1215043) (Fig. 3.9A), 20 SNPs were found within 200kb, which was including the SNP peak (Fig. 3.9B). The LD block shows that SNPs marker was not in the same haploblock with other SNPs around the peak (Fig. 3.9B). The peak was not located in the genic region based on IRRI Rice SNP-Seek Database (Mansueto et al., 2016). Therefore, I did not do deep analysis in chromosome 4.

Seven significant SNPs were classified in chromosome 10 (Fig. 3.10A), these SNPs were block into two clusters (Fig. 3.10B). The first block was 122kb distant, which consisted of 5 SNPs marker, and the second block had two significant SNPs (Fig. 3.10B). Five SNPs in block 1 made 99 rice accessions into four distinct haplotypes (Fig. 3.10C, Table 3.4). Haplotype 1 included 89 accessions which the SES 20 DAT about 5.6, haplotype 3 also had the SES 20 DAT about 4.5 (one accession). The opposite was for haplotype 2 (8 accessions) and haplotype 4 (one accession) which SES 20 DAT was around 9.0 (Fig. 3.10C, Table 3.4). Two SNPs in block two made 99 rice accessions into two distinct haplotypes. Haplotype 1 represented for 91 accessions which the mean of SES 20 DAT about 5.7, while the SES 20 DAT of haplotype 2 was about 8.1 for eight accessions (Fig. 3.10D, Table 3.4).

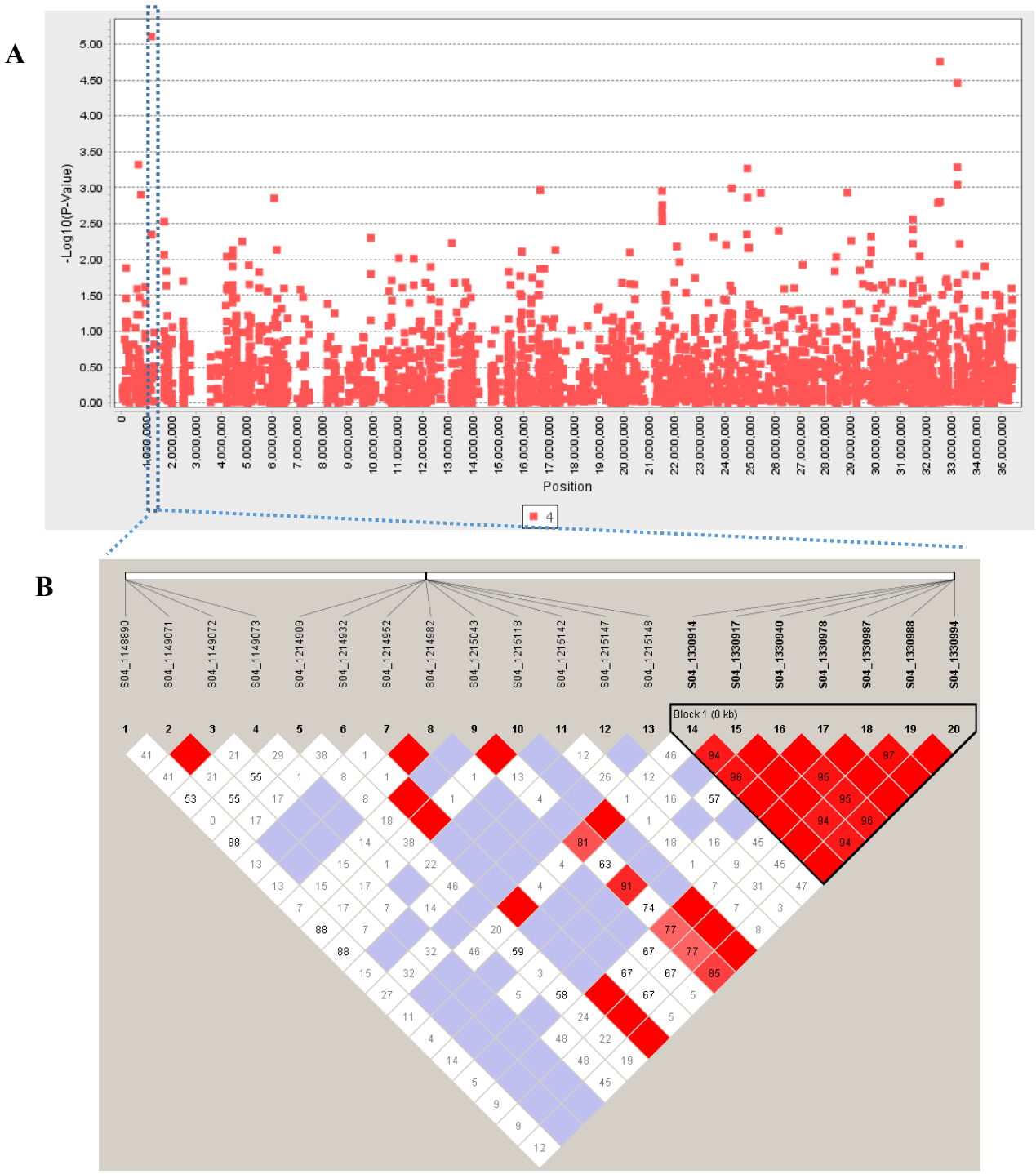


Figure 3.9. Analyses of the peak for salinity traits on chromosome 04

(A) Manhattan plot for SES score, (B) LD block of the peak and other SNP within 200kb

Table 3.4. Haplotype information of significant SNPs on chr10

Haplotype	Block 1							Block 2			
	SNP position					Number of accessions	SES 20 DAT	SNP position		Number of accessions	SES 20 DAT
	5805380	5847722	5847733	5847780	5928011			8601860	8601916		
Hap1	G	C	C	G	T	89	5.6±0.24	C	A	91	5.7±0.24
Hap2	A	T	T	A	C	8	8.3±0.58	T	G	8	8.1±0.59
Hap3	A	C	C	G	T	1	4.5±0.0				
Hap4	G	T	T	A	T	1	9.0±0.0				
Total						99				99	

SES score and shoot dry weight ratio were measured for 97 accessions. One accession in Hap1 and one accession in Hap2 of block 1 were not germinate. Two accessions in Hap1 of block 2 were not germinate. Hap: Haplotype

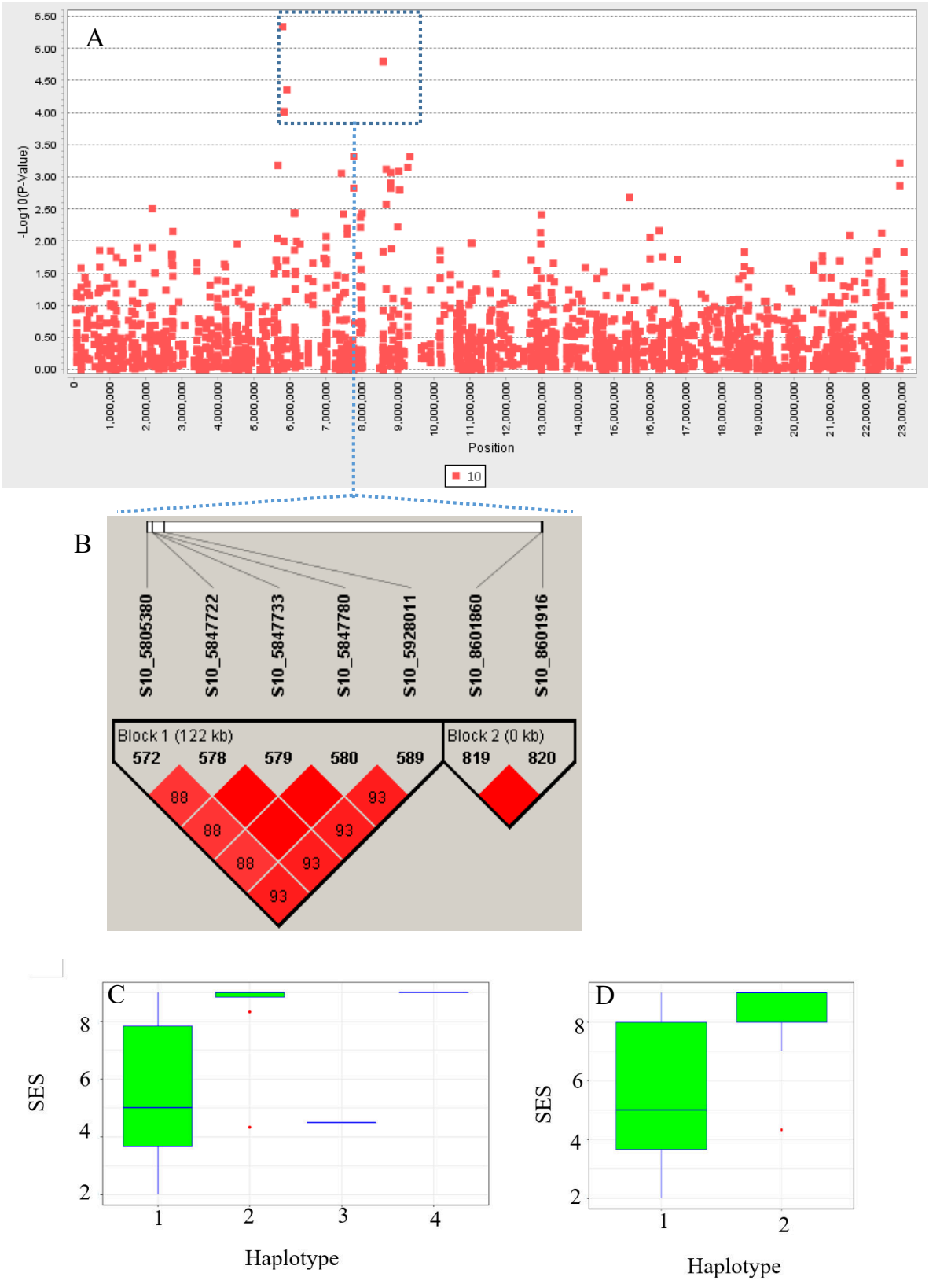


Figure 3.10. Analyses of significant SNP including the SNP peak for salinity traits on chromosome 10

(A) Manhattan plot for SES score on chr10, (B) LD block of significant SNPs on chr10, (C) SES score of four haplotypes in block 1, (D) SES score of two haplotypes in block 2

Table 3.5: Candidate gene of significant SNPs from GWAS and linkage disequilibrium (LD) in chromosome 10

No.	Gene name	Start position	End position	Function
1	LOC_Os10g10530	5,816,937	5,817,345	Expressed protein
2	LOC_Os10g10540	5,821,225	5,828,147	Cysteine-rich repeat secretory protein precursor, putative, expressed
3	LOC_Os10g10550	5,831,375	5,832,212	Hypothetical protein
4	LOC_Os10g10560	5,833,606	5,834,262	Invertase/pectin methylesterase inhibitor family protein, putative, expressed
5	LOC_Os10g10570	5,837,774	5,838,500	Expressed protein
6	LOC_Os10g10580	5,841,678	5,843,555	Expressed protein
7	LOC_Os10g10590	5,850,651	5,852,513	Retrotransposon protein, putative, Ty3-gypsy subclass
8	LOC_Os10g10600	5,852,781	5,857,202	Retrotransposon protein, putative, Ty3-gypsy subclass, expressed
9	LOC_Os10g10610	5,862,796	5,864,638	Retrotransposon, putative, centromere-specific
10	LOC_Os10g10620	5,866,542	5,867,341	Invertase/Pectin methylesterase inhibitor family protein, putative, expressed
11	LOC_Os10g10630	5,874,240	5,874,905	Plant invertase/pectin methylesterase inhibitor domain containing protein, expressed
12	LOC_Os10g10640	5,876,073	5,881,058	Retrotransposon protein, putative, Ty3-gypsy subclass, expressed
13	LOC_Os10g10650	5,884,750	5,885,184	Retrotransposon protein, putative, unclassified
14	LOC_Os10g10660	5,886,846	5,889,653	Retrotransposon protein, putative, Ty3-gypsy subclass
15	LOC_Os10g10664	5,891,518	5,892,006	Hypothetical protein
16	LOC_Os10g10680	5,898,544	5,903,245	Transposon protein, putative, CACTA, En/Spm sub-class, expressed
17	LOC_Os10g10700	5,905,926	5,908,571	Invertase/pectin methylesterase inhibitor family protein, putative%2C expressed
18	LOC_Os10g10720	5,916,507	5,921,052	Retrotransposon protein, putative, unclassified, expressed

Using seven significant SNPs, including the peak of significant SNP to analyze the linkage disequilibrium and haplotype. According the block of LD, 18 genes were classified.

Based on Rice SNP-seek database (Mansueto et al., 2016), I found 18 possible candidate genes located in haploblock 1 within 122kb region. The predicted functions of the genes were retrotransposon protein (8 genes), transposon protein (1 gene), invertase/pectin methylesterase inhibitor family protein (4 genes), cysteine-rich repeat secretory protein (1 gene), expressed protein (3 genes), and hypothetical protein (2 genes) (Table 3.5).

GWAS for four salinity traits of 99 MDI rice accessions

I also used 37,643 SNPs to perform GWAS for four salinity traits of 99 rice accessions. The results illustrate that plant height elongation ratio (Fig. 3.11A), shoot dry weight (Fig. 3.11C) and root dry weight (Fig. 3.11D) were highly correlated with genotype of 99 rice accessions. Whereas, root length elongation was not strongly correlated with any genomic region based on Manhattan plot (Fig. 3.11B).

Related to the correlation between SES 20 DAT and phenotypic traits (Fig. 3.4), I focused on the Manhattan plot of shoot dry weight for deep analysis because the Pearson correlation efficiency was highest among four phenotypic traits (Fig. 3.4). The result shows that nine chromosomes (26 significant SNPs) were highly correlated with shoot dry weight at the threshold –log(p-value) ≥4.0, namely Chr01, Chr02, Chr03, Chr05, Chr06, Chr07, Chr08, Chr10, and Chr12, respectively (Fig. 3.12, Table 3.6).

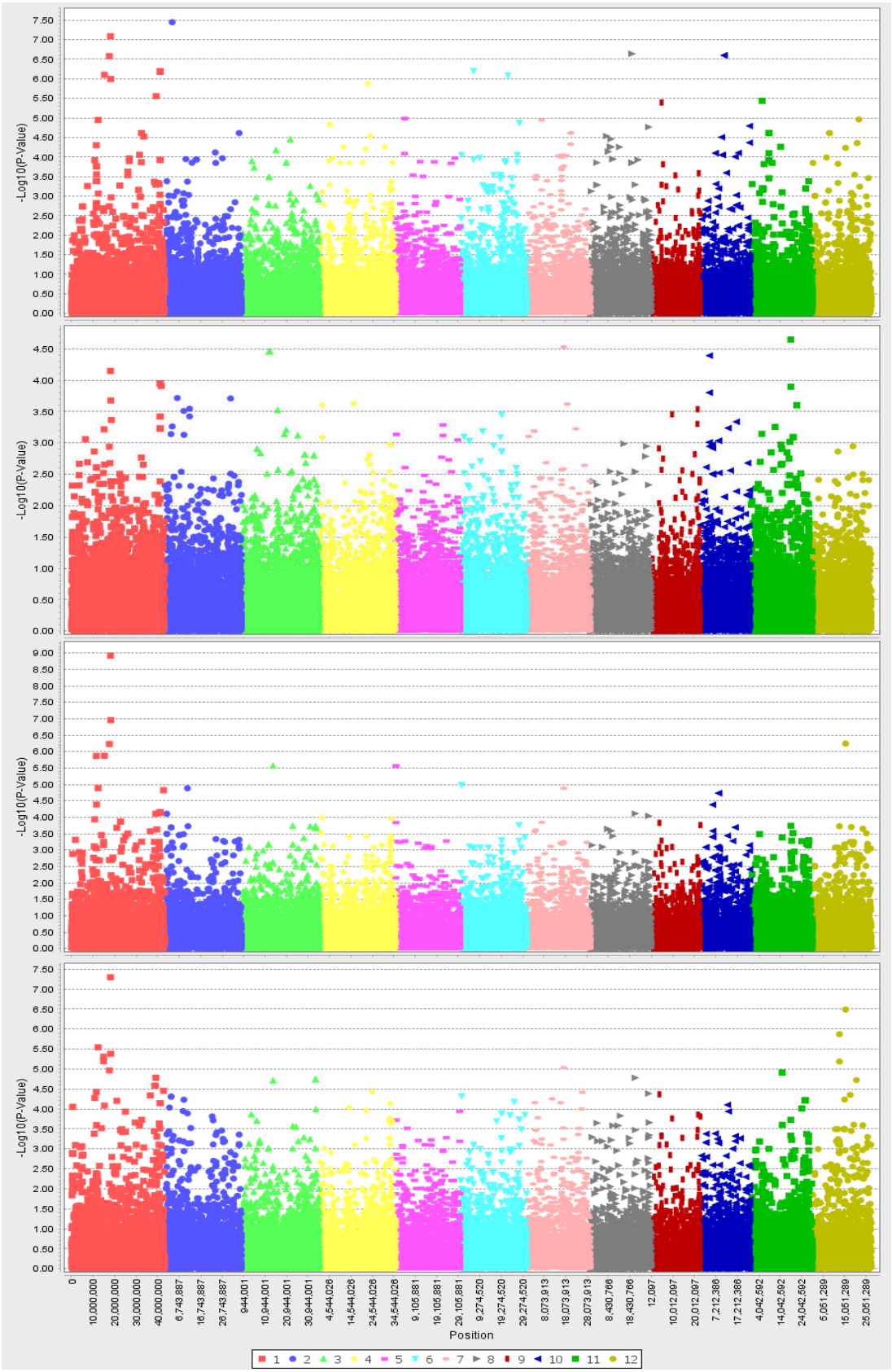


Figure 3.11. GWAS for salinity traits in 100 mM NaCl at 20 DAT. *(A) Manhattan plot for plant height elongation ratio, (B) Manhattan plot for root length elongation ratio, (C) Manhattan plot for shoot dry weight ratio and (D) Manhattan plot for root dry weight ratio.*

Table 3.6. Significant SNPs for shoot dry weight

No.	Marker	Chr	Pos	Marker F	p-value	q-value
1	S01_11466366	1	11466366	15.7343	1.38E-06	0.00866
2	S01_11519675	1	11519675	11.3184	4.12E-05	0.0872
3	S01_12323483	1	12323483	21.2792	1.29E-05	0.03822
4	S01_15190150	1	15190150	15.7601	1.35E-06	0.00866
5	S01_17479548	1	17479548	16.894	5.89E-07	0.00554
6	S01_18015212	1	18015212	26.0318	1.20E-09	4.52E-05
7	S01_18151907	1	18151907	19.2347	1.11E-07	0.00209
8	S01_39153652	1	39153652	10.5281	7.79E-05	0.11895
9	S01_41153286	1	41153286	17.3323	7.12E-05	0.11895
10	S01_41153301	1	41153301	17.3323	7.12E-05	0.11895
11	S01_41153313	1	41153313	17.3323	7.12E-05	0.11895
12	S01_41153328	1	41153328	17.3323	7.12E-05	0.11895
13	S01_42765409	1	42765409	12.5848	1.52E-05	0.04087
14	S02_1011301	2	1011301	10.5107	7.90E-05	0.11895
15	S02_10629393	2	10629393	12.7672	1.31E-05	0.03822
16	S03_14677220	3	14677220	14.8248	2.72E-06	0.01154
17	S05_140370	5	140370	25.0127	2.76E-06	0.01154
18	S05_140371	5	140371	25.0127	2.76E-06	0.01154
19	S06_817638	6	817638	13.0633	1.04E-05	0.03822
20	S07_17260496	7	17260496	12.7625	1.32E-05	0.03822
21	S08_20886757	8	20886757	10.5281	7.79E-05	0.11895
22	S08_27066970	8	27066970	10.3254	9.18E-05	0.13291
23	S10_5805380	10	5805380	11.3057	4.17E-05	0.0872
24	S10_8601860	10	8601860	20.434	1.85E-05	0.04352
25	S10_8601916	10	8601916	20.434	1.85E-05	0.04352
26	S12_15517392	12	15517392	16.9307	5.74E-07	0.00554

Significant SNPS at the level of q-value ≤0.05

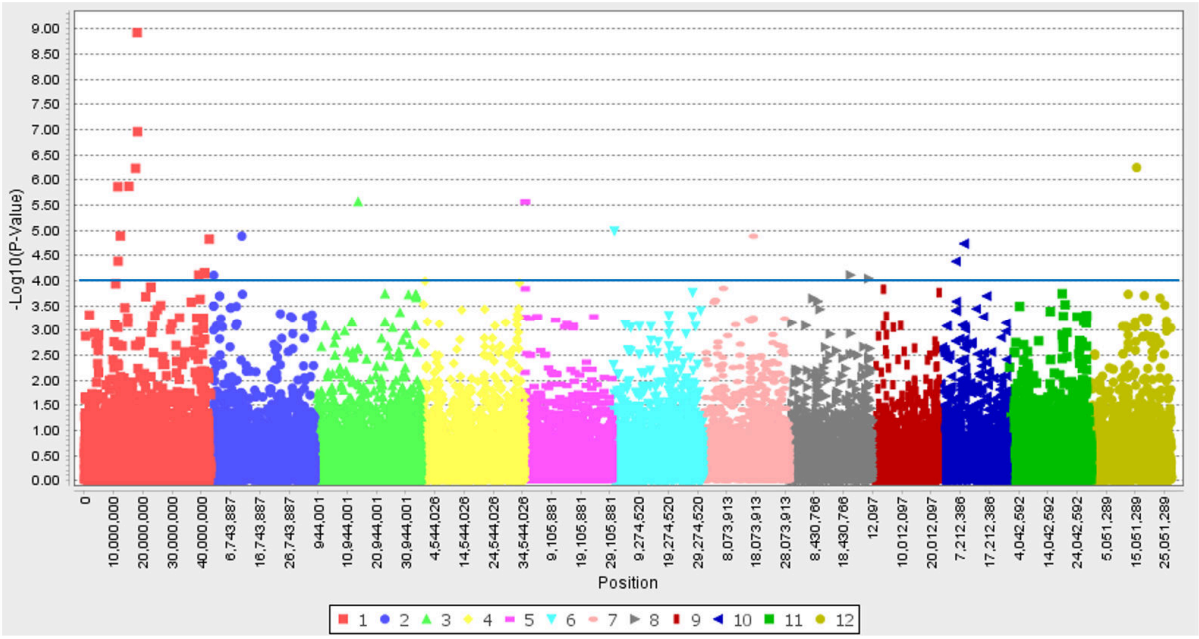


Figure 3.12. GWAS for shoot dry weight in 100 mM NaCl at 20 DAT

Manhattan plot for SES 20 DAT. Using 37,643 SNPs from imputation of 99 accessions and their shoot dry weight to perform GWAS. Chromosomes 01 was highly significant correlated with SES Shoot dry weight. SNP marker at position 18,015,212bp in chromosome 01 was at the peak.

In addition, the SNP peak marker in chr01 was S01_18015212, which was the same position for SES 20 DAT (Fig 3.12). At this position, the shoot dry weight ratio of C allele was lowest, at about 0.29, followed by the A allele at about 0.48, and the highest ratio was for the heterozygous (about 0.50) (Fig. 3.8B, Table 3.3). Using q-value to adjust the significant SNP for GWAS, the objective of this method was to conform whether significant SNPs from GWAS analysis was significant or not at the q-value. SNP markers can be considered as significant when the q-value ≤0.05. Finally, 16 SNPs in 8 chromosomes were identified by q-value (Table 3.6). The largest number of significant SNP was in chromosome 1 (7 SNPs), followed by chromosome 5 and 10 (2 SNPs for each), while five chromosomes (chromosome 2, 3, 6, 7 and 12) had only one SNP (Table 3.6).

3.4. Discussions

Plant growth is influenced by salt stress (Thu et al., 2017). In salt stress, water uptake, root length, shoot length, the fresh weight of shoot and root, dry weight of shoot and root, and some elements were reduced (Gungor et al., 2019). One of the reasons for growth reduction is that root elongation was deduction because of Na^+ could be replaced Ca^{2+} and it caused the rigidity of cell wall (Byrt et al., 2018). In the present study, I conducted screening of salt tolerance hydroponically using 97 rice accessions at seedling stage under 100 mM NaCl stress. Variations of salinity tolerance among these accessions were large. Five accessions had highly tolerance in 100 mM NaCl, namely, Doc Phung (known as salinity tolerant varieties in Vietnam (The et al., 2018)), Trang Phieu, Nang Quot Bien, Mot Bui Lun and Lun Can Trang. Four salinity phenotypic traits were highly and negatively correlated with SES score at 20 DAT, which has been also observed in the previous study (Sexcion et al., 2009). Moreover, I found some accessions showing low SES score (tolerant to salt condition) had the relative root elongation more than 1.0, indicating that NaCl stimulated root elongation in these accessions.

Manhattan plot between phenotype and genotype indicated that three chromosomes had correlated to salinity tolerance as Chr01, Chr04, and Chr10,

respectively. Especially, in present study, chromosome 1 was the best chromosomes for salinity traits, and SNP marker on chromosome 01 (S01_18015212) was the best SNP for identifying the SES score and salinity traits. Using SNP marker S01_18015212 for identifying gene in IRRI Rice SNP-Seek Database (Mansueto et al., 2016) and RAP DB (Sakai et al., 2013), I found the gene LOC_Os01g32830 which the function was similar to LrgB-like family protein, with highly correlation to SES score 20 DAT, plant height elongation, root length elongation, shoot dry weight and root dry weight. SNP marker S01_18015212 is located in 3' UTR of the gene, it may be affected on expressed protein because the 3' UTR have majority role on mRNA regulation of localization, translation, and stability (Sun et al., 2017). Moreover, SNP marker S01_18015212 is located in three QTLs of salinity tolerance (QTAROqtl-195, QTAROqtl-196, QTAROqtl-197) (Mansueto et al., 2016), which the function is Na^+ uptake, K^+ concentration and the ratio between Na^+ and K^+, respectively. Therefore, SNP marker S01_18015212 can be used for identifying salinity tolerance.

Regarding on chromosome 10, I identified 18 genes that they explained for SES score of MDI rice population. In addition, four genes, LOC_Os10g10560, LOC_Os10g10620, LOC_Os10g10630 and LOC_Os10g10700 have the function of invertase/pectin methylesterase inhibitor family protein which are known for cell wall of plant (Nguyen et al., 2016, Byrt et al., 2018, Wang et al., 2019). In salt condition, Na^+ interference in pectin cross-linking could reduce the stabilizing influence of pectin in the cell wall and then it is likely to trigger mechanisms to rigidify the cell wall. This mechanism may contribute to reduce root elongation in saline soil (Byrt et al., 2018). Therefore, these genes are related for rice growth and development (Nguyen et al., 2016). I also found eight retrotransposon protein genes, accounted for 44.4% of identified genes in this region. Interestingly, the function of retrotransposon gene is related to stress factors; it will be activating in stress and environmental factors (Todorovska, 2014).

Four phenotypic traits were highly negative correlation with SES 20 DAT, but the shoot dry weight ratio was strongest correlation, therefore, I focused on performing GWAS of shoot dry weight ratio. Manhattan plot of shoot dry weight show that SNP marker S01_18015212 was the strongest evidence for this trait because p-value was

$1.20E^{-09}$ and q-value was $4.52E^{-05}$. The present study shows that S01_18015212 was useful for identifying salinity tolerant varieties.

3.5. Conclusions

I determined five accessions with high tolerance to 100 mM NaCl in 20 DAT, and 47 accessions were tolerant.

The phenotypic traits of 97 accessions was affected by salt condition, therefore, the ratio between NaCl treatment and control treatment on plant height elongation, root length elongation, shoot dry weight and root dry weight were almost lower than 0.5.

The SES score increased with the increased of day treatment.

Genome-wide association study of traits related to seedling stage salinity tolerance was highly significant in chr01 and chr10. I found five genes that high correlation to salinity tolerance were LOC_Os01g32830, LOC_Os10g10560, LOC_Os10g10620, LOC_Os10g10630 and LOC_Os10g10700.

Chapter 4. Summary

Rice (*Oryza sativa* L.) farming is the principal agricultural activity in many countries, particularly, South and Southeast Asia countries. Rice is a staple food because of directly feeding more people than any other cereals. Nearly half of the world's population and 90% of Asian people rely on rice every day. In Vietnam, rice produced in the Mekong Delta contributes to 54% of the total rice productivity and 90% of the total export volume of Vietnam. Currently, about 2,000 samples of local and improved rice varieties in the Mekong Delta have been collected and conserved at Mekong Delta Development Research Institute (MDI) - Can Tho University (Viet Nam) as the genetic material for salinity tolerance rice breeding.

In this study, I used 99 *indica* rice accessions collected from 10 provinces in the coastal region of Mekong Delta in Vietnam to evaluate the genomic structure and screen salinity tolerance at seedling stage. Moreover, I would like to identify the relationship between genotype and phenotype on salt-traits by using SNPs from RAD-seq and whole genome sequence.

1. Profiling SNP and nucleotide diversity to characterize Mekong Delta rice landraces in Southeast Asian populations

After double digest restriction-site associated DNA sequencing of the 99 rice accessions and subsequent filtering; 2,301 SNPs were generated and used in subsequent analyses. Within the 3K dataset, MDI landraces fell into an Ind3 cluster consisting of accessions from Southeast Asian countries, while MDI improved varieties were grouped in cluster Ind1B. A principal component analysis suggested that geographical location strongly affects phylogenetic relationships of Southeast Asian rice accessions, and the MDI landraces were placed into a Vietnam+Cambodia group. The genomic distribution of π-values representing the nucleotide diversity of each population also reflected these phylogenetic relationships and suggested a genetic adaptation to Southeast Asian locations. To display the characteristics of local populations, I constructed a genomic map as a simple profile representing low π-value regions. My

simple profiling using low π-value genomic regions was able to reveal regional characteristics of rice genomes and should be useful for identifying local rice populations.

2. The genome-wide association study of salinity responses

A total of 99 accessions were phenotyped and genotyped against salinity tolerance at seedling stage in 100 mM sodium chloride (NaCl) under hydroponic conditions. After growing the accessions for 20 days under NaCl stress, I found 5 highly tolerant, 47 moderately tolerant, 17 sensitive and 28 highly sensitive accessions. The standard evaluation system (SES) was applied to evaluate the salinity tolerance/sensitivity for the examined accession. The strong correlations were observed between the SESs and relative elongation values of the shoot height, root length, shoot and root dry weights under NaCl treatment at 20 days after treatment (DAT). Based on genome wide association study (GWAS), I found three genomic regions in chr01, chr04, and chr10, which located QTLs related to salinity traits. Particularly, chr01 and chr10 were highly associated with SES score at all 4 treatment periods; 5 DAT, 10 DAT, 15 DAT and 20 DAT. I also found Doc Phung cultivar that was extremely tolerant to salt stress at 20 DAT. This cultivar is useful for salinity tolerance breeding programs.

Supplement

Supplementary Table 1. SES score and salinity traits of MDI's rice accessions

No.	Accessions code	SES 5 DAT	SES 10 DAT	SES 15 DAT	SES 20 DAT	Plant height elongation ratio	Root length elongation ratio	Shoot weight ratio	Root weight ratio
1	MDI-1	3.50	4.50	5.50	9.00	0.12	0.21	0.14	0.16
2	MDI-2	3.33	5.00	5.67	8.33	0.10	0.20	0.16	0.19
3	MDI-3	2.33	3.67	4.00	5.00	0.16	1.17	0.27	0.31
4	MDI-4	3.00	3.33	4.00	5.00	0.15	0.37	0.29	0.29
5	MDI-5	2.00	3.00	3.33	3.67	0.17	1.17	0.31	0.33
6	MDI-6	2.00	2.33	3.33	3.67	0.08	0.96	0.37	0.43
7	MDI-7	2.50	3.50	4.50	9.00	0.03	0.34	0.23	0.21
8	MDI-8	2.00	2.00	3.00	3.50	0.36	1.14	0.54	0.50
9	MDI-9	2.50	3.50	4.50	7.50	0.19	0.17	0.21	0.21
10	MDI-11	3.33	3.67	3.67	4.00	0.21	0.80	0.28	0.25
11	MDI-12	2.00	2.00	3.00	3.00	0.40	0.69	0.44	0.36
12	MDI-13	2.00	2.33	2.33	2.33	0.50	1.20	0.45	0.51
13	MDI-14	2.00	2.33	2.67	2.67	0.77	1.09	0.50	0.51
14	MDI-15	2.33	3.00	3.00	3.67	0.24	0.79	0.35	0.36
15	MDI-16	2.00	3.33	4.00	5.00	0.35	1.03	0.40	0.40
16	MDI-17	2.00	2.00	2.33	2.67	0.35	1.07	0.56	0.71
17	MDI-18	3.50	4.50	6.00	8.00	0.04	0.22	0.24	0.19
18	MDI-19	3.00	3.33	3.33	3.33	0.53	0.69	0.75	0.86
19	MDI-21	2.00	2.00	2.00	2.00	0.59	1.08	0.83	0.95
20	MDI-22	3.00	3.33	3.33	3.33	0.18	0.71	0.36	0.42
21	MDI-23	4.00	5.33	7.00	9.00	0.18	0.02	0.17	0.17
22	MDI-24	3.33	3.33	3.33	3.33	0.29	1.68	0.56	0.73
23	MDI-25	3.00	4.50	5.00	7.50	0.12	0.36	0.26	0.31
24	MDI-26	2.33	3.00	3.67	4.00	0.18	0.44	0.43	0.56
25	MDI-27	3.50	4.50	5.00	9.00	0.08	0.09	0.19	0.24
26	MDI-28	3.00	4.00	4.67	8.33	0.11	0.09	0.32	0.28
27	MDI-29	2.50	4.50	6.50	7.50	0.12	0.36	0.29	0.37
28	MDI-30	3.50	5.00	7.50	9.00	0.17	0.06	0.13	0.14
29	MDI-31	2.33	3.33	3.33	3.67	0.42	0.58	0.40	0.40
30	MDI-41	2.67	4.00	4.33	7.33	0.06	0.42	0.18	0.17
31	MDI-42	2.00	2.67	3.00	3.67	0.14	1.02	0.39	0.46
32	MDI-43	3.33	6.00	7.00	9.00	0.09	0.22	0.16	0.15
33	MDI-44	3.00	3.50	4.00	4.00	0.33	0.72	0.50	0.46
34	MDI-49	3.00	4.50	5.00	8.00	0.16	0.15	0.21	0.18
35	MDI-50	2.33	3.67	4.00	5.33	0.13	0.31	0.27	0.25
36	MDI-52	2.00	3.00	3.00	3.33	0.16	0.98	0.39	0.43
37	MDI-53	2.33	3.33	4.00	5.67	0.19	0.85	0.27	0.28
38	MDI-54	3.00	3.33	4.00	4.67	0.31	0.46	0.31	0.30
39	MDI-55	2.33	3.33	4.00	4.67	0.23	0.94	0.39	0.44
40	MDI-56	2.00	3.00	3.00	4.00	0.38	0.57	0.46	0.48
41	MDI-57	2.00	3.00	3.33	4.33	0.21	0.18	0.45	0.39

No.	Accessions code	SES 5 DAT	SES 10 DAT	SES 15 DAT	SES 20 DAT	Plant height elongation ratio	Root length elongation ratio	Shoot weight ratio	Root weight ratio
42	MDI-58	2.33	3.00	3.67	4.00	0.20	0.72	0.47	0.49
43	MDI-59	2.00	2.00	2.67	3.33	0.35	1.35	0.47	0.65
44	MDI-60	2.00	2.50	3.00	3.50	0.18	0.06	0.49	0.36
45	MDI-61	2.33	3.33	3.33	3.67	0.18	0.16	0.43	0.32
46	MDI-62	2.33	3.00	3.33	4.33	0.16	1.03	0.40	0.45
47	MDI-63	2.33	3.67	3.67	4.33	0.20	0.64	0.32	0.41
48	MDI-64	2.33	4.00	4.00	4.33	0.11	0.16	0.30	0.29
49	MDI-65	2.67	3.00	3.33	3.33	0.24	0.47	0.39	0.39
50	MDI-66	2.33	3.00	3.33	3.67	0.24	0.59	0.42	0.47
51	MDI-67	3.50	4.50	4.50	9.00	0.19	0.20	0.26	0.28
52	MDI-68	2.67	3.33	4.00	7.67	0.13	0.31	0.32	0.30
53	MDI-69	2.50	3.50	4.00	5.00	0.22	0.97	0.41	0.39
54	MDI-70	3.00	5.33	6.00	8.67	0.13	0.22	0.15	0.15
55	MDI-71	3.33	4.67	4.67	8.33	0.05	0.21	0.31	0.36
56	MDI-72	3.33	4.67	5.00	8.33	0.05	0.14	0.24	0.27
57	MDI-75	3.00	3.67	4.33	4.67	0.13	0.03	0.35	0.27
58	MDI-77	4.00	5.00	5.00	5.00	0.16	0.62	0.29	0.31
59	MDI-82	3.00	4.00	4.33	4.67	0.19	0.81	0.39	0.39
60	MDI-83	3.00	4.00	4.00	4.50	0.13	0.80	0.31	0.32
61	MDI-85	3.00	4.00	4.00	4.00	0.16	0.53	0.32	0.30
62	MDI-86	3.00	4.00	4.33	5.00	0.10	0.76	0.30	0.37
63	MDI-87	4.00	5.00	5.67	8.33	0.04	0.01	0.12	0.12
64	MDI-88	2.67	3.00	3.00	3.33	0.40	1.28	0.61	0.77
65	MDI-89	5.00	6.67	8.33	9.00	0.03	0.01	0.08	0.13
66	MDI-90	2.00	2.33	2.67	3.33	0.07	0.22	0.51	0.50
67	MDI-91	4.50	7.50	9.00	9.00	0.05	0.05	0.10	0.12
68	MDI-92	3.50	4.00	6.00	7.00	0.02	0.37	0.19	0.22
69	MDI-95	3.00	3.67	5.33	7.00	0.04	0.38	0.32	0.28
70	MDI-96	4.33	6.33	7.33	9.00	0.06	0.13	0.12	0.11
71	MDI-97	3.00	4.00	4.50	5.00	0.28	0.19	0.33	0.30
72	MDI-98	4.67	7.67	8.67	9.00	0.08	0.01	0.15	0.21
73	MDI-99	2.67	3.67	4.33	7.33	0.07	0.40	0.43	0.40
74	MDI-100	3.00	3.67	4.67	6.67	0.08	0.39	0.34	0.37
75	MDI-101	3.67	4.67	6.67	7.67	0.17	0.05	0.25	0.30
76	MDI-102	3.00	4.00	4.67	6.67	0.12	0.70	0.42	0.43
77	MDI-104	3.50	3.50	3.50	4.00	0.23	0.48	0.48	0.40
78	MDI-105	2.50	3.00	4.00	4.50	0.12	1.03	0.33	0.36
79	MDI-106	3.33	5.67	7.00	9.00	0.04	0.05	0.17	0.19
80	MDI-108	4.67	7.00	8.33	9.00	0.03	0.01	0.09	0.13
81	MDI-109	3.50	4.50	5.50	8.50	0.05	0.48	0.31	0.30
82	MDI-111	4.50	7.00	9.00	9.00	0.02	0.00	0.11	0.17
83	MDI-112	4.67	7.00	8.67	9.00	0.06	0.00	0.11	0.18
84	MDI-118	2.00	2.67	3.33	3.33	0.25	0.65	0.54	0.68
85	MDI-119	3.67	5.00	6.00	9.00	0.13	0.16	0.17	0.21

No.	Accessions code	SES 5 DAT	SES 10 DAT	SES 15 DAT	SES 20 DAT	Plant height elongation ratio	Root length elongation ratio	Shoot weight ratio	Root weight ratio
86	MDI-120	3.33	5.33	8.00	9.00	0.01	0.21	0.14	0.17
87	MDI-121	4.67	6.67	8.67	9.00	0.03	0.02	0.15	0.20
88	MDI-122	3.33	4.67	5.00	8.67	0.10	0.17	0.18	0.21
89	MDI-123	2.50	3.00	3.50	3.50	0.29	0.84	0.57	0.88
90	MDI-125	2.67	3.67	4.00	5.67	0.19	0.05	0.41	0.38
91	MDI-126	2.00	3.00	4.00	5.00	0.11	0.21	0.35	0.41
92	MDI-127	3.33	4.00	4.33	6.33	0.11	0.48	0.32	0.38
93	MDI-128	3.50	5.00	7.00	9.00	0.07	0.11	0.16	0.21
94	MDI-129	2.67	3.00	3.33	3.67	0.17	0.98	0.39	0.46
95	MDI-130	3.00	3.50	4.00	5.00	0.08	0.78	0.26	0.28
96	MDI-131	2.67	3.00	3.33	4.33	0.36	0.57	0.51	0.64
97	MDI-133	3.67	4.00	4.67	8.33	0.06	0.49	0.21	0.28

SES: System evaluation standard for salt injury

DAT: Days after treatment

Supplementary Table 2. Identifying SNPs in gen LOC_Os01g32830

No.	Marker	Chromosome	Position	p-value
1	S01_18015061	1	18015061	0.14295
2	S01_18015064	1	18015064	0.04291
3	S01_18015102	1	18015102	0.04567
4	S01_18015198	1	18015198	0.56886
5	S01_18015202	1	18015202	0.27401
6	S01_18015212	1	18015212	8.77E-08
7	S01_18015432	1	18015432	0.99389
8	S01_18015476	1	18015476	0.49492
9	S01_18015478	1	18015478	0.49388

Position: 18013675..18017027 (3353bp) (- strand)

Acknowledgments

Firstly, I would like to extend my sincere gratitude to my supervisor Professor Yuji Kishima, Plant Breeding Laboratory - Hokkaido University, for his instructive advice and useful suggestions on my studying from researching to finishing doctoral thesis. Beside my advisor, I would like to profusely thank Dr. Maria Stefanie Dwiyanti and Associate Professor Toshihiro Watanabe (Hokkaido University) for their insightful advice and comments on my paper and doctoral thesis. I aslo thank Dr. Itsuro Takamure (Hokkaido University) and Assistant Professor Yohei Koide (Hokkaido University) for their massive supports.

My sincere acknowledgments also go to all members of Plant Breeding Laboratory, especially Mr. Sakaguchi, Mr. Yamamori, Ms. Nozomi Saito for their great kindness and supports during my doctoral course. I am thankful to all the staffs of Hokkaido University I have met, and to my friends at Hokkaido University in the past three years.

I would like to express my thanks to the board of directors and my colleagues of Mekong Delta Development Research Institute – Can Tho University for helping and supporting prepare materials and this study. Moreover, I also thanks to the rectorate board of Can Tho University and the administrative of ODA project give me the opportunity to improve my education in Japan.

I express my profound thanks to Vietnamese Government Scholarship Council for providing the scholarship to me during my doctoral course.

Last but not least, I sincerely appreciate my family: my parents, my wife and my son for supporting me throughout my study and my life in general.

References

Acquaah G, 2015. Conventional Plant Breeding Principles and Techniques. *Springer, Cham* In: Al-Khayri J., Jain S., Johnson D. (eds) Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools.

Ahmadi N, Negrao S, Katsantonis D*, et al.*, 2011. Targeted association analysis identified japonica rice varieties achieving Na+/K+ homeostasis without the allelic make-up of the salt tolerant indica variety Nona Bokra. *Theoretical and Applied Genetics* 123, 881-95.

Alexander DH, Lange K, 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12.

Alexander DH, Novembre J, Lange K, 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19, 1655-64.

Ali ML, Mcclung AM, Jia MH, Kimball JA, Mccouch SR, Eizenga GC, 2011. A rice diversity panel evaluated for genetic and agro-morphological diversity between subpopulations and its geographic distribution. *Crop Science* 51, 2021-35.

Ashikari M, Sasaki A, Ueguchi-Tanaka M*, et al.*, 2002. Loss-of-function of a rice gibberellin biosynthetic gene, GA20 oxidase (GA20ox-2), led to the rice 'green revolution'. *Breeding Science* 52, 143-50.

Ashraf M, Akram N, Foolad M, 2012. Marker-assisted selection in plant breeding for salinity tolerance. In. *Plant Salt Tolerance.* Springer, 305-33.

Baird NA, Etter PD, Atwood TS*, et al.*, 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *Plos One* 3.

Batayeva D, Labaco B, Ye C*, et al.*, 2018. Genome-wide association study of seedling stage salinity tolerance in temperate japonica rice germplasm. *BMC Genetics* 19, 2.

Bolger AM, Lohse M, Usadel B, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-20.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES, 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633-5.

Browning BL, Zhou Y, Browning SR, 2018. A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics* 103, 338-48.

Byrt CS, Munns R, Burton RA, Gilliham M, Wege S, 2018. Root cell wall solutions for crop plants in saline soils. *Plant Science* 269, 47-55.

Calpe C, 2006. Rice international commodity profile. *Rome: Food and Agricultural Organization of the United Nations*.

Chang T-T, 2003. Origin, domestication, and diversification. *Rice: origin, history, technology, and production* 3, 3.

Chen KM, Luan MB, Xiong HP*, et al.*, 2018. Genome-wide association study discovered favorable single nucleotide polymorphisms and candidate genes associated with ramet number in ramie (*Boehmeria nivea* L.). *BMC Plant Biology* 18.

Clauss K, Ottinger M, Leinenkugel P, Kuenzer C, 2018. Estimating rice production in the Mekong Delta, Vietnam, utilizing time series of Sentinel-1 SAR data. *International Journal of Applied Earth Observation and Geoinformation* 73, 574-85.

Dabney A, Storey JD, Warnes G, 2004. Q-value estimation for false discovery rate control. *Medicine* 344, 48.

De Leon TB, Linscombe S, Subudhi PK, 2016. Molecular Dissection of Seedling Salinity Tolerance in Rice (Oryza sativa L.) Using a High-Density GBS-Based SNP Linkage Map. *Rice* 9, 52.

De Leon TB, Linscombe S, Subudhi PK, 2017. Identification and validation of QTLs for seedling salinity tolerance in introgression lines of a salt tolerant rice landrace 'Pokkali'. *PLoS ONE* 12, e0175361.

Depristo MA, Banks E, Poplin R*, et al.*, 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43, 491-+.

Emon RM, Islam MM, Halder J, Fan Y, 2015. Genetic diversity and association mapping for salinity tolerance in Bangladeshi rice landraces. *The Crop Journal* 3, 440-4.

Evenson RE, Gollin D, 1994. Genetic resources, international organizations, and rice varietal improvement. In.: Center Discussion Paper.

Flowers T, Koyama M, Flowers S, Sudhakar C, Singh K, Yeo A, 2000. QTL: their place in engineering tolerance of rice to salinity. *Journal of Experimental Botany* 51, 99-106.

Foti C, Khah EM, Pavli OI, 2019. Germination profiling of lentil genotypes subjected to salinity stress. *Plant Biology* 21, 480-6.

Frouin J, Languillaume A, Mas J*, et al.*, 2018. Tolerance to mild salinity stress in japonica rice: A genome-wide association mapping study highlights calcium signaling and metabolism genes. *PLoS ONE* 13, e0190964.

Furukawa T, Maekawa M, Oki T*, et al.*, 2007. The Rc and Rd genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant Journal* 49, 91-102.

Genc Y, Oldach K, Verbyla AP*, et al.*, 2010. Sodium exclusion QTL associated with improved seedling growth in bread wheat under salinity stress. *Theoretical and Applied Genetics* 121, 877-94.

Gepts P, 2006. Plant genetic resources conservation and utilization: The accomplishments and future of a societal insurance policy. *Crop Science* 46, 2278-92.

Ghomi K, Rabiei B, Sabouri H, Sabouri A, 2013. Mapping QTLs for Traits Related to Salinity Tolerance at Seedling Stage of Rice (*Oryza sativa* L.): An Agrigenomics Study of an Iranian Rice Population. *Omics-a Journal of Integrative Biology* 17, 242-51.

Gilliham M, Able JA, Roy SJ, 2017. Translating knowledge about abiotic stress tolerance to breeding programmes. *Plant Journal* 90, 898-917.

Gungor H, Cikili Y, Dumlupinar Z, 2019. EVALUATION OF MORPHO-PHYSIOLOGICAL TRAITS OF TURKISH RICE GENOTYPES IN RESPONSE TO SALT STRESS UNDER IN VITRO CONDITIONS. *Journal of Animal and Plant Sciences* 29, 556-67.

Hariadi YC, Nurhayati AY, Soeparjono S, Arif I, 2015. Screening Six Varieties of Rice (Oryzasativa) for Salinity Tolerance. *Procedia Environmental Sciences* 28, 78-87.

Hermisson J, Wagner GP, 2004. The population genetic theory of hidden variation and genetic robustness. *Genetics* 168, 2271-84.

Hirano HY, Sano Y, 1991. Molecular characterization of the Waxy locus of rice (*Oryza sativa*). *Plant and Cell Physiology* 32, 989-97.

Hoang TML, Moghaddam L, Williams B, Khanna H, Dale J, Mundree SG, 2015. Development of salinity tolerance in rice by constitutive-overexpression of genes involved in the regulation of programmed cell death. *Frontiers in Plant Science* 6.

Hsu JM, Lee YCG, Yu CTR, Huang CYF, 2004. Fbx7 functions in the SCF complex regulating Cdk1-cyclin B-phosphorylated hepatoma up-regulated protein (HURP) proteolysis by a proline-rich region. *Journal of Biological Chemistry* 279, 32592-602.

Imran QM, Kamran M, Rehman SU*, et al.*, 2016. GA Mediated OsZAT-12 expression improves salt resistance of rice. *International Journal of Agriculture and Biology* 18, 330-6.

Inta A, Balslev H, Gustafsson MHG*, et al.*, 2016. Genetic diversity patterns of rice (*Oryza sativa* L.) landraces after migration by Tai Lue and Akha between China and Thailand. *Genetic Resources and Crop Evolution* 63, 845-58.

Irri, 2014. *Standard evaluation for rice (SES)*. Los Baños, Philippines.

Islam M, Hassan L, Salam M, Collard B, Singh R, Gregorio G, 2011. QTL mapping for salinity tolerance at seedling stage in rice. *Emirates Journal of Food and Agriculture*, 137-46.

Komiya R, Ikegami A, Tamaki S, Yokoi S, Shimamoto K, 2008. Hd3a and RFT1 are essential for flowering in rice. *Development* 135, 767-74.

Konishi S, Izawa T, Lin SY, *et al.*, 2006. An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392-6.

Kontgis C, Schneider A, Ozdogan M, *et al.*, 2019. Climate change impacts on rice productivity in the Mekong River Delta. *Applied Geography* 102, 71-83.

Kovach MJ, Calingacion MN, Fitzgerald MA, Mccouch SR, 2009. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences of the United States of America* 106, 14444-9.

Koyama ML, Levesley A, Koebner RM, Flowers TJ, Yeo AR, 2001. Quantitative trait loci for component physiological traits determining salt tolerance in rice. *Plant physiology* 125, 406-22.

Krishnamurthy P, Ranathunge K, Nayak S, Schreiber L, Mathew MK, 2011. Root apoplastic barriers block Na+ transport to shoots in rice (*Oryza sativa* L.). *Journal of Experimental Botany* 62, 4215-28.

Kumari S, Panjabi Nee Sabharwal V, Kushwaha HR, Sopory SK, Singla-Pareek SL, Pareek A, 2008. Transcriptome map for seedling stage specific salinity stress response indicates a specific set of genes as candidate for saline tolerance in *Oryza sativa* L. *Functional & Integrative Genomics* 9, 109.

Langmead B, Salzberg SL, 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357-U54.

Lee K-S, Choi W-Y, Ko J-C, Kim T-S, Gregorio GB, 2003. Salinity tolerance of japonica and indica rice (Oryza sativa L.) at the seedling stage. *Planta* 216, 1043-6.

Lee S, Ahn J, Cha Y, *et al.*, 2007. Mapping QTLs related to salinity tolerance of rice at the young seedling stage. *Plant Breeding* 126, 43-6.

Lin H, Zhu M, Yano M, *et al.*, 2004. QTLs for Na+ and K+ uptake of the shoots and roots controlling rice salt tolerance. *Theoretical and Applied Genetics* 108, 253-60.

Linh Le H, Linh TH, Xuan TD, Ham Le H, Ismail AM, Khanh TD, 2012. Molecular Breeding to Improve Salt Tolerance of Rice (Oryza sativa L.) in the Red River Delta of Vietnam. *Int J Plant Genomics* 2012, 949038.

Mansueto L, Fuentes RR, Borja FN, *et al.*, 2016. Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic acids research* 45, D1075-D81.

Martin ML, Busconi L, 2000. Membrane localization of a rice calcium-dependent protein kinase (CDPK) is mediated by myristoylation and palmitoylation. *Plant Journal* 24, 429-35.

Mckenna A, Hanna M, Banks E, *et al.*, 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20, 1297-303.

Mitra J, 2001. Genetics and genetic improvement of drought resistance in crop plants. *Current science*, 758-63.

Mittler R, Blumwald E, 2010. Genetic engineering for modern agriculture: challenges and perspectives. *Annual review of plant biology* 61, 443-62.

Munns R, Tester M, 2008. Mechanisms of salinity tolerance. *Annu Rev Plant Biol* 59, 651-81.

Mursyidin DH, Purnomo P, Sumardi I, Daryono BS, 2018. Molecular diversity of tidal swamp rice (*Oryza sativa* L.) in South Kalimantan, Indonesia. *Diversity-Basel* 10.

Muthayya S, Sugimoto JD, Montgomery S, Maberly GF, 2014. An overview of global rice production, supply, trade, and consumption. *Annals of the new york Academy of Sciences* 1324, 7-14.

Nakhoda B, Leung H, Mendioro MS, Mohammadi-Nejad G, Ismail AM, 2012. Isolation, characterization, and field evaluation of rice (Oryza sativa L., Var. IR64) mutants with altered responses to salt stress. *Field Crops Research* 127, 191-202.

Nations OFaaOOTU, 2015. Agricultural outlook 2015-2024. *OECD Publishing, Paris*.

Nguyen HP, Jeong HY, Kim H, Kim YC, Lee C, 2016. Molecular and biochemical characterization of rice pectin methylesterase inhibitors (OsPMEIs). *Plant Physiology and Biochemistry* 101, 105-12.

Nhung TT, Vo PV, Nghi VV, Bang HQ, 2019. Salt intrusion adaptation measures for sustainable agricultural development under climate change effects: A case of Ca Mau Peninsula, Vietnam. *Climate Risk Management* 23, 88-100.

Novembre J, 2014. Variations on a Common STRUCTURE: New algorithms for a valuable model. *Genetics* 197, 809-11.

Penjor T, Mimura T, Kotoda N*, et al.*, 2016. RAD-Seq analysis of typical and minor Citrus accessions, including Bhutanese varieties. *Breeding Science* 66, 797-807.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE, 2012. Double Digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One* 7.

Phan NT, Trinh LT, Rho MY*, et al.*, 2019. Identification of loci associated with fruit traits using genome-wide single nucleotide polymorphisms in a core collection of tomato (*Solanum lycopersicum* L.). *Scientia Horticulturae* 243, 567-74.

Phung NTP, Mai CD, Mournet P*, et al.*, 2014b. Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. *Bmc Plant Biology* 14, 371.

Platten JD, Egdane JA, Ismail AM, 2013. Salinity tolerance, Na+ exclusion and allele mining of HKT1;5 in Oryza sativa and O-glaberrima: many sources, many genes, one mechanism? *BMC Plant Biology* 13.

Puram VRR, Ontoy J, Linscombe S, Subudhi PK, 2017. Genetic Dissection of Seedling Stage Salinity Tolerance in Rice Using Introgression Lines of a Salt Tolerant Landrace Nona Bokra. *Journal of Heredity* 108, 658-70.

Pusadee T, Prom-U-Thai C, Yimyam N, Jamjod S, Rerkasem B, 2017. Phenotypic and genetic diversity of local perilla (*Perilla frutescens* (L.) Britt.) from Northern Thailand. *Economic Botany* 71, 175-87.

Quan R, Wang J, Hui J*, et al.*, 2018. Improvement of Salt Tolerance Using Wild Rice Genes. *Frontiers in Plant Science* 8, 2269-.

Ray S, Satya P, 2014. Next generation sequencing technologies for next generation plant breeding. *Frontiers in Plant Science* 5.

Reddy INBL, Kim B-K, Yoon I-S, Kim K-H, Kwon T-R, 2017. Salt Tolerance in Rice: Focus on Mechanisms and Approaches. *Rice Science* 24, 123-44.

Reich D, Price AL, Patterson N, 2008. Principal component analysis of genetic data. *Nature Genetics* 40, 491-2.

Rellosa MC, Reano RA, Capilit GLS*, et al.*, 2014. The 3,000 rice genomes project. *Gigascience* 3.

Rivera RM, Solis SHD, 2019. Traits association of rice phytogenetic resources collections in Los Palacios. *Avances* 21, 22-31.

Sakai H, Lee SS, Tanaka T*, et al.*, 2013. Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant and Cell Physiology* 54, e6-e.

Samberg LH, Shennan C, Zavaleta E, 2013. Farmer seed exchange and crop diversity in a changing agricultural landscape in the southern highlands of Ethiopia. *Human Ecology* 41, 477-85.

Sexcion FSH, Egdane JA, Ismail AM, Dionisio-Sese ML, 2009. Morpho-physiological Traits Associated with Tolerance of Salinity During Seedling Stage in Rice (Oryza sativa L.). *Philippine Journal of Crop Science* 34, 27-37.

Shah SM, Arif M, Aslam K, Shabir G, Thomson MJ, 2016. Genetic diversity analysis of Pakistan rice (*Oryza sativa*) germplasm using multiplexed single nucleotide polymorphism markers. *Genetic Resources and Crop Evolution* 63, 1113-26.

Shavrukov Y, Suchecki R, Eliby S, Abugalieva A, Kenebayev S, Langridge P, 2014. Application of next-generation sequencing technology to study genetic diversity and identify unique SNP markers in bread wheat from Kazakhstan. *Bmc Plant Biology* 14.

Shi Y, Gao L, Wu Z*, et al.*, 2017. Genome-wide association study of salt tolerance at the seed germination stage in rice. *BMC Plant Biology* 17, 92.

Singh R, Gregorio GB, Jain R, 2007. QTL mapping for salinity tolerance in rice. *Physiology and Molecular Biology of Plants* 13, 87.

Sun HX, Li Y, Niu QW, Chua NH, 2017. Dehydration stress extends mRNA 3' untranslated regions with noncoding RNA functions in Arabidopsis. *Genome Res* 27, 1427-36.

Tang S, Zhang Y, Zeng L, Luo L, Zhong Y, Geng Y, 2010. Assessment of genetic diversity and relationships of upland rice accessions from southwest China using microsatellite markers. *Plant Biosystems* 144, 85-92.

Thanh NC, Duong BD, Hien TV, Minh NH, Singh M, 2013. Study on rice production and marketing of farmers in Mekong Delta. *Omonrice*, 13.

The HV, Hien NQ, Cuong PB, 2018. Evaluating salt tolerance of twenty traditional rice varieties from Vietnam. *Tạp chí Khoa học công nghệ và Thực phẩm* 14, 3-11.

Thomson MJ, De Ocampo M, Egdane J*, et al.*, 2010. Characterizing the Saltol quantitative trait locus for salinity tolerance in rice. *Rice* 3, 148.

Thu TTP, Yasui H, Yamakawa T, 2017. Effects of salt stress on plant growth characteristics and mineral content in diverse rice genotypes. *Soil Science and Plant Nutrition* 63, 264-73.

Tian L, Tan L, Liu F, Cai H, Sun C, 2011. Identification of quantitative trait loci associated with salt tolerance at seedling stage from Oryza rufipogon. *Journal of Genetics and Genomics* 38, 593-601.

Todorovska E, 2014. Retrotransposons and their Role in Plant—Genome Evolution. *Biotechnology & Biotechnological Equipment* 21, 294-305.

Tu M, Lu BR, Zhu Y, Wang Y, 2007. Abundant within-varietal genetic diversity in rice germplasm from Yunnan Province of China revealed by SSR fingerprints. *Biochemical Genetics* 45, 789-801.

Vaughan DA, Lu B-R, Tomooka N, 2008. The evolving story of rice evolution. *Plant Science* 174, 394-408.

Wang JC, Guan YJ, Wang Y*, et al.*, 2014. A strategy for finding the optimal scale of plant core collection based on Monte Carlo simulation. *Scientific World Journal*.

Wang W, Mauleon R, Hu Z*, et al.*, 2018. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43-9.

Wang WL, Xu Y, Chen TX*, et al.*, 2019. Regulatory mechanisms underlying the maintenance of homeostasis in Pyropia haitanensis under hypersaline stress conditions. *Science of the Total Environment* 662, 168-79.

Yamanaka S, Jatoi SA, San-San-Yi, Kothari SL, Tin-Htut, Watanabe KN, 2011. Genetic diversity of Myanmar rice and their implementation on management methods. *African Journal of Biotechnology* 10, 1290-8.

Yang HA, Tao Y, Zheng ZQ, Li CD, Sweetingham MW, Howieson JG, 2012. Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. *BMC Genomics* 13.

Yousef EaA, Muller T, Borner A, Schmid KJ, 2018. Comparative analysis of genetic diversity and differentiation of cauliflower (*Brassica oleracea* var. botrytis) accessions from two ex situ genebanks. *Plos One* 13.

Yu J, Zao W, He Q, Kim T-S, Park Y-J, 2017. Genome-wide association study and gene set analysis for understanding candidate genes involved in salt tolerance at the rice seedling stage. *Molecular Genetics and Genomics* 292, 1391-403.

Zhao K, Wright M, Kimball J*, et al.*, 2010. Genomic diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome. *Plos One* 5, e10780.

Zheng HL, Zhao HW, Liu HL, Wang JG, Zou DT, 2015. QTL analysis of Na+ and K+ concentrations in shoots and roots under NaCl stress based on linkage and association analysis in japonica rice. *Euphytica* 201, 109-21.

Zhu MY, Wang YY, Zhu YY, Lu BR, 2004. Estimating genetic diversity of rice landraces from Yunnan by SSR assay and its implication for conservation. *Acta Botanica Sinica* 46, 1458-67.