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; Phan et al., 2019; Wang et al., 2014), 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 and Wagner, 2004; Penjor et al., 2016; Phung et al., 2014; Rivera and Solis, 2019; Yousef et al., 2018; Zhu et al., 2004).

Genetic markers such as simple sequence repeats (SSR) and Restriction Fragment Length Polymorphism (RFLP) are popular in genetic diversity analyses because

**ABSTRACT** Recent analyses using single nucleotide polymorphism (SNP) are a feasible mean for local collections which potentially possess useful, but not large, genetic variations. Genomic sequences of more than 3000 accessions released by the International Rice Research Institute (IRRI) can be used to characterize various local rice (*Oryza sativa*) populations. The aim of this study was to develop a method to facilitate genomic characterization of local rice populations. We mainly used 99 *indica* rice accessions (81 landraces and 18 improved varieties) from the Mekong Delta Development Research Institute (MDI). We obtained 2301 SNPs after a genomic sequencing analysis of the 99 rice accessions and subsequent filtering. Within the IRRI’s dataset, the landraces fell into a cluster consisting of accessions from Southeast Asian countries (Ind3 cluster), and the MDI improved varieties were grouped in a cluster containing IRRI improved varieties (Ind1B cluster). A principal component analysis suggested that geographical location strongly affects phylogenetic relationships, and the MDI landraces were placed into a Vietnam+Cambodia group. To detect the nucleotide diversity within a population, \( p \)-value is commonly used. We think that whole genome distribution of \( p \)-values representing the nucleotide diversity of each population can be used to characterize local populations. Our simple profiling using low \( p \)-value genomic regions was able to reveal regional characteristics of rice genomes and should be useful for identifying local rice populations.

**CORE IDEAS**

- Single nucleotide polymorphism (SNP) analyses are a powerful tool to examine structure of local rice population.
- 3000 dataset of IRRI facilitates SNP profiling of Southeast Asian rice populations.
- Mekong Delta population is featured by comparisons with the other populations.
- The low \( p \)-value SNPs well-profile unique genetic regions in their genomes.

**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; Phan et al., 2019; Wang et al., 2014), 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 and Wagner, 2004; Penjor et al., 2016; Phung et al., 2014; Rivera and Solis, 2019; Yousef et al., 2018; Zhu et al., 2004).

Genetic markers such as simple sequence repeats (SSR) and Restriction Fragment Length Polymorphism (RFLP) are popular in genetic diversity analyses because

**Abbreviations:** Chr, chromosome; ddRAD-seq, double digest restriction-site associated DNA sequencing; IRRI, International Rice Research Institute; MAF, minor allele frequency; MDI, Mekong Delta Development Research Institute; NGS, next-generation sequencing; PCA, Principal component analysis; SNP, Single nucleotide polymorphism.
the procedures are quite simple, the costs are affordable for labs with small budgets, and the analyses 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: (i) it is biallelic, (ii) it is abundant in genome, and (iii) 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 genotyping methods based on NGS and high-throughput genotyping are being applied to a variety of crops (Chen et al., 2018; Ray and Satya, 2014; Shah et al., 2016; Shavrukov et al., 2014; Yang et al., 2012). Because of their affordability, these newly improved methods can be exploited for use in small research projects with limited budgets. In particular, many currently untouchable bio-resource collections preserved in local institutions are now ready for characterization (Mursyidin et al., 2018; Tú et al., 2007; Yamanaka et al., 2011). 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 3010 rice (Oryza sativa) 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 3000 dataset contains over 29 million SNPs and 2.4 million small indels. A genomic analysis classified Asian cultivated rice into nine subgroups and three admixtures, with 90,000 structural variations identified within and among populations. The nine subgroups are four indica (XI) clusters (XI-1A from East Asia, XI-1B of improved varieties from various origins, XI-2 from South Asia, and XI-3 from Southeast Asia); three japonica (GJ) clusters (primarily East Asian temperate, GJ-temp; Southeast Asian subtropical, GJ-stbrp; and Southeast Asian tropical, GJ-trp); and two groups from South Asia: Aus, Boro, and Rayada group (cA), and Sadri-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 to contribute 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 Vietnam which is responsible for 19% of the global rice market (Clauss et al., 2018). In the 1970s, the improved line such as IR5 started being introduced from IRRI to Mekong Delta, followed by dwarf varieties represented by IR8, IR36, and MTL30. Recent varieties with high productivity, like OM576, OM2451, and IR50404, have been used in the rice production of 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, we 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, are henceforth referred to as the MDI set (Fig. 1A; Supplemental Table S1). Using SNP genotype data generated by double digest restriction-site associated DNA sequencing (ddRAD-seq) of the MDI set, we attempted to establish a general procedure for the evaluation of the genomic structure of local bio-resources. To achieve this objective, we used the rice 3000 dataset as a reference, and performed three different genomic analyses to assess the population structure, phylogenetic relationships, and diversity of the MDI set. We 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, we developed a simple method to profile local rice populations using regions of low genetic diversity. The resulting profile is available for identification and comparison with other rice populations.

MATERIALS AND METHODS

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 (Fig. 1A; Supplemental Table S1). 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 ecosystems. Our materials did not include some Mekong Delta leading varieties such as OM576, OM2514, and IR50404, although we considered that the pedigrees for the 18 improved varieties used in this study may share the similar genetic structures.

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 to 10 d after germination. From approximately 100 mg of fresh young leaves, DNA was extracted using a DNeasy Plant Mini kit (Qiagen, Hilden, Germany) following the manufacturer’s protocol. The quantity of DNA was measured using a Thermo Scientific NanoDrop 2000 spectrophotometer (Fisher Scientific, Hampton, NH), and DNA quality was checked by 1% (w/v) agarose gel electrophoresis.
Each DNA sample was digested with EcoRI–BglII enzymes and ligated to restriction-site associated DNA sequencing (RAD-seq) adaptor sets (Baird et al., 2008; Peterson et al., 2012). The DNA-adaptor sets were pooled into a library and sequenced on a HiSeq2500 system using the 100-bp paired-end sequencing method (Illuminia, San Diego, CA).

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 with 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 and Salzberg, 2012) available in Galaxy (https://usegalaxy.org.au/, accessed Dec. 2018). The reads were further filtered using Picard (http://broadinstitute.github.io/picard, accessed Dec. 2018), and the alignments were adjusted around indels using the IndelRealigner tool provided in Genome Analysis Toolkit v2.8 (McKenna et al., 2010). Single nucleotide polymorphism 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 (Supplemental Table S2).

Phylogenetic and Population Structure Analyses

For genetic diversity analyses, the initial SNP dataset was filtered according to the following parameters: missing call rate, 100%; minor allele frequency (MAF), 0.05; and heterozygosity rate, 0.02. The resulting dataset of 2301 high-quality SNPs was compared with the SNPs of the 1789 indica accessions in the 3000 SNP dataset (29 million SNPs) which yielded a combined SNP dataset of 2301 SNPs for genetic diversity analyses. Principal component analysis (PCA) of the MDI set and indica accessions was performed in TASSEL v5.2.43 (Bradbury et al., 2007), and the results were plotted using R. Phylogenetic analysis was performed 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/, accessed 14 Sept. 2017).
Population structure analysis was performed using ADMIXTURE v1.23 (Alexander and Lange, 2011; Alexander et al., 2009). 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 (Fig. 2).

Nucleotide Diversity Analysis
Using 2301 SNPs, we performed an analysis of the nucleotide diversity of the 99 MDI rice accessions from the Mekong Delta and 412 Ind3 rice accessions from the 3000 dataset (Supplemental Fig. S1). The 412 Ind3 rice accessions are from eight Southeast Asian countries. The $\pi$-value is an index indicating nucleotide diversity at SNP site; the low $\pi$-value expresses the low nucleotide differences among the samples. Computation of $\pi$-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.

RESULTS AND DISCUSSION

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 (Supplemental Table S1). The materials were collected from 10 provinces, all located in coastal regions of the Mekong Delta in Vietnam (Fig. 1A; Supplemental Table S1). 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 2301 SNPs distributed across 12 chromosomes (Supplemental Fig. S1). Using the 2301 SNPs, phylogenetic relationships of the 99-accssion MDI set were examined by the neighbor-joining method (Fig. 1B). In the resulting phylogenetic tree, MDI improved varieties were clearly distinguished from MDI landraces (Fig. 1B). The MDI set can be divided into seven groups based on provincial origin. 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), but no clear clustering based on provincial origin was observed. Thus, we 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. 1B) 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. 1B, 2).

Integration of the MDI Set with the 3000 Dataset
The 2301 SNPs from the MDI set were analyzed along with the corresponding SNPs in the rice 3000 dataset published by the IRRI (Wang et al., 2018). To reveal the relationships of the MDI accessions to those of the 3000 set, we 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. 3A). The MDI landraces fell into cluster Ind3 of the 3000 set, which consisted of accessions from Southeast Asian countries (Laikat 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, we conducted a 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 3000 set (Wang et al., 2018; Fig. 3B; Supplemental Figure S2). The PCA clustered the 493 accessions into four groups (Fig. 3B): 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 suggest that genetic and geographical relationships within the Ind3 subgroup are correlated. In fact, several previous studies also showed the correlations between the genome constitutions and geographical locations in the rice accessions distributed in Southeast Asia (Laikat Ali et al., 2011; Inta et al., 2016; Pusadee et al., 2017; Tang et al., 2010, Wang et al., 2018). Exceptionally, the group containing the accessions from the Philippines and Myanmar was geographically discontinuous (Fig. 3B).

The PCA indicated that the MDI landraces mostly belonged to the Vietnam+Cambodia group (Fig. 3B), and the neighbor-joining phylogenetic analysis distributed the MDI landraces into two clusters (Fig. 3C). 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. 3C; Supplemental Fig. S2). The MDI landraces falling into the Cambodian group were collected from provinces close to Cambodia, such as An Giang and Kien Giang (Fig. 1, 3C).

Nucleotide Diversities of MDI Landraces and Improved Varieties

When evaluating genetic resources, genetic diversity of the resources is an important factor, which can be measured by nucleotide diversity. The $\pi$-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 2301 SNPs, we 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, we 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 \( \pi \)-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 (Supplemental Fig. S3). Low- and high-diversity regions were defined as those with \( \pi \)-values below 0.05 or over 0.45, respectively (Supplemental Tables S3, S4). In the case where the \( \pi \)-values used below 0.1 or over 0.40, the corresponding regions increased more than two folds of the cases with the \( \pi \)-values below 0.05 or over 0.45. Therefore, we selected the regions with \( \pi \)-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 (Supplemental Table S3), while high diversity was correspondingly indicated in 30 and 52 regions (Supplemental Table S4). Figure 4 is a graphical representation of the two nucleotide diversity profiles with different colors indicating \( \pi \)-value levels at each five-SNP interval throughout the genome. As shown in Fig. 4, low- and high-diversity regions were located in different genomic positions between the two groups. The fact that low-diversity regions were rarely shared between MDI landraces and MDI improved varieties implies that MDI landraces and MDI improved varieties are far from each other in terms of genetic distance, as shown in phylogenetic trees (Fig. 1, 3A). Therefore, we assumed that these genomic regions in MDI landraces might have contributed little to the genomic structures of MDI improved varieties.

**Comparisons of Genetic Diversities among Landraces in Southeast Asian Countries**

Because MDI landraces were grouped with Vietnamese and Cambodian rice accessions (Fig. 3B), we assessed whether they shared similar genetic diversity profiles. Using the same 2301 SNPs as in the above analysis, we estimated genetic diversities of 459 regions (every five SNPs) with 412 accessions from eight countries in Southeast Asia (Supplemental Fig. S4). 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 (\( r = 0.58 \)), and Vietnam (\( r = 0.69 \)) among the other countries.

Fig. 4. Circular genomic profiles of nucleotide diversities of Mekong Delta Development Research Institute (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 (\( \pi \)-values) are color-coded as follows: red, high diversity; blue, low diversity.
= 0.64), and Thailand (r = 0.46); in contrast, the genome-wide nucleotide diversity profile of MDI landraces differed from that of Malaysian (r = 0.26) and Indonesian (r = 0.25) accessions (Table 1). These results are consistent with the groupings uncovered by the PCA (Fig. 3B).

The above-mentioned geographical groups uncovered by the SNP analysis were correlated with locations of low and high genetic diversity in the genome (Fig. 5). When we 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. 5), we observed a geographical correlation pattern that resembled that which was obtained from the PCA (Fig. 3B; Supplemental Fig. S2).

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. 5; Supplemental Fig. S4; Supplemental Table S3). 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

Table 1. Pairwise correlation matrix of genome-wide nucleotide diversity of Ind3 accessions by country.

<table>
<thead>
<tr>
<th></th>
<th>MDI† landraces</th>
<th>Vietnam</th>
<th>Cambodia</th>
<th>Thailand</th>
<th>Myanmar</th>
<th>Philippines</th>
<th>Laos</th>
<th>Malaysia</th>
<th>Indonesia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vietnam</td>
<td>0.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambodia</td>
<td>0.64</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>0.46</td>
<td>0.51</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myanmar</td>
<td>0.41</td>
<td>0.47</td>
<td>0.61</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philippines</td>
<td>0.34</td>
<td>0.43</td>
<td>0.53</td>
<td>0.63</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laos</td>
<td>0.30</td>
<td>0.35</td>
<td>0.50</td>
<td>0.76</td>
<td>0.62</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td>0.26</td>
<td>0.33</td>
<td>0.43</td>
<td>0.59</td>
<td>0.48</td>
<td>0.52</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>0.25</td>
<td>0.36</td>
<td>0.47</td>
<td>0.58</td>
<td>0.51</td>
<td>0.55</td>
<td>0.38</td>
<td>0.76</td>
<td></td>
</tr>
</tbody>
</table>

† MDI, Mekong Delta Development Research Institute.

---

**Diagram from outer to inner:**
- MDI landraces
- Vietnam
- Cambodia
- Myanmar
- Laos
- Thailand
- Philippines
- Malaysia
- Indonesia

**π-values**
- >0.45
- 0.4
- 0.35
- 0.3
- 0.25
- 0.2
- 0.15
- 0.1
- 0.05
- 0

\(\triangleright\) : Common high diversity
\(\triangleleft\) : Common low diversity

---

Fig. 5. Circular profiles of nucleotide diversities of Mekong Delta Development Research Institute (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 \(\pi\)-values. Red and blue arrows respectively indicate high- and low-diversity regions shared by six of eight groups representing different Southeast Asian countries.
and window sizes of five SNPs (Supplemental Fig. S4). In MDI landraces, but not MDI improved varieties, all three regions were marked by low diversity (Fig. 4; Supplemental Fig. S3; Supplemental Table S3). 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 (Sakai et al., 2013) and Rice SNP-Seek (Manuel et al., 2017) 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; Supplemental Table S5). The known gene RICE FLOWERING LOCUS T1 (RFT1; Komiyama et al., 2008) is present in the low-diversity region of Chr 6 (Supplemental Table S5). In contrast, no known genes in these databases could be aligned with the low-diversity region of Chr 7 (Supplemental Table S5). The genes present in these low-diversity regions may be involved in specific adaptation in Southeast Asian countries; however, Sh4 (Martin and Busconi, 2000), qSH1 (Konishi et al., 2006), Sdi1 (Ashikari et al., 2002), Wx (Hirano and Sano, 1991), Badh2.1 (Kovach et al., 2009), and Rc (Furukawa et al., 2007), identified as low genetic-diversity genes in analyses of the 3000 set, were not included in these three regions (Wang et al., 2018).

Aside from shared low-diversity regions, we 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. 5; Supplemental Table S4). This high genetic diversity suggests the presence of positive selection. Although we cannot currently link the high-diversity regions to such positive selection, the fact that the high-diversity region found at 14.1 to 14.7 Mb on Chr 12 was shared by all Southeast Asian groups (Supplemental Table S4).

In addition to shared low- and high-diversity regions, we respectively found four and three low- and high genetic diversity regions specific to MDI accessions (Supplemental Tables S3, S4). 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 1000 genes are distributed in these regions (Supplemental Table S3, S4), we 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.

**Simple Profiling for Low Genetic Diversity**

As shown in Fig. 5 and Supplemental Fig. S4, we characterized the genetic diversities of MDI landraces and rice populations from eight Southeast Asian countries. We especially focused on low-diversity regions (π <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. 6, we 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 2301 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 3000 set can be used to compare rice populations and determine their genetic inter-relationships.

**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 (Fig. 1, 2). Genomic structural analyses using the 3000 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 comprised of modern improved varieties such as IR64 (Fig. 3A). 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 3000 set and sorted them into four groups (Fig. 3B; Table 1) possibly related to geography. The MDI landraces were closely related to the Vietnam+Cambodia group but were far from the Indonesia+Malaysia group (Fig. 3B, 3C). These relationships were also clearly reflected in the pattern of nucleotide (genetic) diversity (Table 1). 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 (Fig. 5). Low π-value regions are particularly useful as unique genomic signatures of local accessions (Fig. 6). The genomic regions with such low π-values may be associated with local adaptations of the landraces. We thus propose that simple profiling of a local rice population using low π-value regions can be used for identification purposes.

**Funding**

This study was funded in part by the Can Tho University Improvement Project VN14-P6, supported by a Japanese ODA loan.

**Data Availability**

The raw sequence data of the 99-accession MDI set generated in this study have been deposited in the DDBJ database (accession number DRA008414).
Conflict of Interest

The authors declare that there is no conflict of interest.

Author Contributions

Nguyen Thanh Tam and Yuji Kishima planned the research. Nguyen Thanh Tam, Maria Stefanie Dwiyanti, and Yuji Kishima designed the research. Nguyen Thanh Tam and Yuji Kishima prepared rice seed samples, and Nguyen Thanh Tam and Maria Stefanie Dwiyanti carried out the experiments. Nguyen Thanh Tam, Maria Stefanie Dwiyanti, and Huynh Ky analyzed the data. Huynh Ky, Huynh Quang Tin, and Le Viet Dung participated in planning. Yohei Koide supervised the materials preparation and data analysis. Atsushi J. Nagano assisted with the ddRAD-seq analysis. Nguyen Thanh Tam, Maria Stefanie Dwiyanti, and Yuji Kishima wrote and improved the manuscript. Maria Stefanie Dwiyanti, Yohei Koide, and Yuji Kishima supervised Nguyen Thanh Tam’s PhD study. All authors have read and approved the manuscript.

Fig. 6. Simple circular profiles of low diversity in Mekong Delta Development Research Institute (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.
We gratefully acknowledge Dr. T. Watanabe and Dr. I. Takamure (Research Faculty of Agriculture, Hokkaido University) for their valuable suggestions concerning this study. We thank Mr. T.H. Phuc, Mr. H.N. Dien, and Mr. N.V. Chanh (Mekong Delta Development Research Institute, Can Tho University) for preparing and maintaining rice seeds for this study. We also thank Mr. S. Sakaguchi for helping with experimental preparations. We thank Edanz Group (www.edanzediting.com/ac) for editing the English text of a draft of this manuscript.

REFERENCES


Shavrukov, Y., R. Suchecki, S. Elby, A. Abuagaleva, S. Kenehayev, and P. Langrudge. 2014. Application of next-generation sequencing technology to study genetic diversity and identify unique SNP markers in bread

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. T. Watanabe and Dr. I. Takamure (Research Faculty of Agriculture, Hokkaido University) for their valuable suggestions concerning this study. We thank Mr. T.H. Phuc, Mr. H.N. Dien, and Mr. N.V. Chanh (Mekong Delta Development Research Institute, Can Tho University) for preparing and maintaining rice seeds for this study. We also thank Mr. S. Sakaguchi for helping with experimental preparations. We thank Edanz Group (www.edanzediting.com/ac) for editing the English text of a draft of this manuscript.


