




## Article

# Estimation of the Genetic Diversity and Population Structure of Thailand's Rice Landraces Using SNP Markers

Wanchana Aesomnuk <sup>1,2</sup>, Siriphat Ruengphayak <sup>3</sup>, Vinitchan Ruanjaichon <sup>4</sup> , Tanee Sreewongchai <sup>5</sup>, Chanate Malumpong <sup>6</sup>, Apichart Vanavichit <sup>3,6</sup>, Theerayut Toojinda <sup>4</sup>, Samart Wanchana <sup>4,\*</sup>  and Siwaret Arikrit <sup>3,6,\*</sup> 

- <sup>1</sup> Center for Agricultural Biotechnology (CAB), Kamphaeng Saen Campus, Kasetsart University, Nakhon Pathom 73140, Thailand; wanchana.a@ku.th
  - <sup>2</sup> Center of Excellence on Agricultural Biotechnology: (AG-BIO/PERDO-CHE), Bangkok 10900, Thailand
  - <sup>3</sup> Rice Science Center, Kamphaeng Saen Campus, Kasetsart University, Nakhon Pathom 73140, Thailand; phatsterster@gmail.com (S.R.); vanavichit@gmail.com (A.V.)
  - <sup>4</sup> National Center for Genetic Engineering and Biotechnology (BIOTEC), 113 Thailand Science Park, Pahonyothin Road, Khlong Nueng, Khlong Luang, Pathum Thani 12120, Thailand; vinitchan.rua@biotec.or.th (V.R.); theerayut@biotec.or.th (T.T.)
  - <sup>5</sup> Department of Agronomy, Faculty of Agriculture, Kasetsart University, Bangkok, Bangkok 10900, Thailand; taneesree@yahoo.com
  - <sup>6</sup> Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kamphaeng Saen Campus, Kasetsart University, Nakhon Pathom 73140, Thailand; agrcnm@ku.ac.th
- \* Correspondence: samart.wan@biotec.or.th (S.W.); siwaret.a@ku.th (S.A.)



**Citation:** Aesomnuk, W.; Ruengphayak, S.; Ruanjaichon, V.; Sreewongchai, T.; Malumpong, C.; Vanavichit, A.; Toojinda, T.; Wanchana, S.; Arikrit, S. Estimation of the Genetic Diversity and Population Structure of Thailand's Rice Landraces Using SNP Markers. *Agronomy* **2021**, *11*, 995. <https://doi.org/10.3390/agronomy11050995>

Academic Editor: Yong-Bao Pan

Received: 19 April 2021

Accepted: 14 May 2021

Published: 17 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Rice is a staple food for more than half of the world's population. Modern rice varieties have been developed for high yield and quality; however, there has been a substantial loss of diversity. A greater number of genetically dynamic landraces could offer valuable and useful genetic resources for rice improvement. In this study, the genetic diversity and population structure of 365 accessions of lowland and upland landraces from four populations from different geographical regions of Thailand were investigated using 75 SNP markers. Clustering analyses using maximum likelihood, Principal Coordinate Analysis (PCoA), and Discriminant Analysis of Principal Components (DAPC) clustered these landraces into two main groups, corresponding to *indica* and *japonica* groups. The *indica* group was further clustered into two subgroups according to the DAPC and STRUCTURE analyses ( $K = 3$ ). The analysis of molecular variance (AMOVA) analysis results revealed that 91% of the variation was distributed among individuals, suggesting a high degree of genetic differentiation among rice accessions within the populations. Pairwise  $F_{ST}$  showed the greatest genetic differentiation between the northeastern and southern populations and the smallest genetic differentiation between the northern and northeastern populations. Isolation-by-distance analysis based on a Mantel test indicated a significant relationship between the genetic distance and geographic distance among the Thai rice landraces. The results from this study provide insight into the genetic diversity of Thai rice germplasm, which will enhance the germplasm characterization, conservation, and utilization in rice genetics and breeding.

**Keywords:** rice (*Oryza sativa*); genetic diversity; population structure; landrace; Thai rice; SNPs

## 1. Introduction

Rice (*Oryza sativa* L.) is one of the most important cereal grains in the world and serves as a staple food for more than half of the world's population [1]. Rice is grown in more than 100 countries, with 90% of the global production from Asia [2]. Research has estimated that about 120,000 distinct rice varieties exist in the world [1]. Southeast Asia is important as the main source of rice germplasm and the source of rice diversity [3]. Asian cultivated rice (*O. sativa* L.) was domesticated from the wild rice species, *O. nivara* and *O. rufipogon* [1], and can be assigned to either the subspecies *indica* or *japonica* depending

on a variety of physiological and morphological characteristics [4]. The subspecies *indica* is primarily grown in the tropical regions, while the subspecies *japonica* is grown in either the subtropical and temperate regions of East Asia [5].

Modern rice varieties have been developed for both high yield and high quality [6]. However, these rice varieties exhibit a loss of diversity, which leads to a wide range of effects regarding adaptations to changing environments [7]. Landraces, on the other hand, have been found to be genetically dynamic and display equilibrium with both the environment and pathogens; thus, they could provide valuable and useful genetic resources for crop improvement [8]. Landrace variation has become interesting knowledge for conservation and utilization [9]. However, landraces have declined in popularity, and a number of landraces are gradually being replaced by improved varieties. There is an urgent need to conserve the landraces that are rapidly declining as high-yield varieties become predominant.

Studies on genetic diversity and population structure are critical for characterizing the genetic relationship among germplasm accessions. Genetic variation among populations and their genetic relationships aid in the conservation and parental selection in crop improvement programs [10]. The identification of populations with a high level of genetic variation will become a valuable resource for broadening the genetic base as this enables the identification of superior alleles for several traits [11]. Numerous techniques can be used to determine rice genetic variations in order to comprehend and use accessible gene bank accessions. Initially, isozyme markers were used to evaluate rice genetic diversity [12]; however, these have now been replaced with DNA markers, such as simple sequence repeats (SSRs), microsatellites [13,14], and single nucleotide polymorphisms (SNPs) [15–18].

Thailand is one of the countries with a significant number of indigenous rice varieties and landraces, which can serve as a valuable genetic resource for future crop improvement to satisfy the ever-increasing demand for food production. Over 17,000 traditional landraces have been conserved at the National Rice Gene Bank collection. To effectively utilize this germplasm, an assessment and classification of the diversity is necessary. While studies on genetic diversity are available for many rice collections around the world [19,20], those for Thai rice germplasm have been conducted on limited sets of accessions [21–25]. According to the previous studies on the genetic structure of Thai rice evaluated using indel and simple sequence repeat (SSR) markers, two distinct groups of rice accessions have been revealed to correspond to *indica* and *japonica* groups [23,24]. However, the genetic structures of lowland and upland landraces from different geographical regions of Thailand are still largely unknown.

In this study, we used a set of SNP markers to investigate the genetic diversity and population structure of a collection of lowland and upland landraces, consisting of 365 accessions obtained from four geographical regions of Thailand. The results from this study provide an insight into the genetic diversity of the Thai rice germplasm, which will enhance germplasm characterization, conservation, and utilization in rice genetics and breeding.

## 2. Materials and Methods

### 2.1. Plant Materials

A collection of 365 accessions of Thai rice landraces (*Oryza sativa* L.) was obtained from the National Rice Gene Bank of Thailand. According to the germplasm registry, this collection is composed of 169 and 196 accessions of upland and lowland rice, respectively, which were obtained from different parts of Thailand. Among the upland rice, there were 46, 38, 40, and 45 accessions from northern (N), northeastern (NE), central (C), and southern (S) populations, respectively. Among the lowland rice, there were 50, 48, 49, and 49 from the N, NE, C, and S populations, respectively (Table S1).

## 2.2. DNA Extraction and SNP Genotyping

The total genomic DNA was extracted from young leaves using a DNA Trap I kit (DNA Technology Laboratory, Thailand). The DNA was quantified using a NanoDrop 8000 Spectrophotometer (Thermo Fisher Scientific, DE, USA) and diluted to a concentration of 5 ng/ $\mu$ L. Genotyping was performed with 102 SNP markers using Kompetitive allele specific PCR (KASP) and the SNPlane genotyping system following standard KASP protocols (LGC Genomics, Teddington, UK). The SNP marker set (102 markers) was a courtesy of the Rice Science Center, Kasetsart University, Thailand (data unpublished).

## 2.3. Data Management and Analysis

The input for all analyses was based on the genotype data from polymorphic SNP markers. The genetic distance was calculated using the Nie's standard distance [26], and the phylogenetic tree was created using MEGA X [27] based on the Maximum-likelihood method [28].

The major allele frequency (MAF), number of alleles per locus ( $N_a$ ), and gene diversity per locus ( $H_e$ ) as well as the polymorphism information content (PIC) value were estimated using PowerMarker version 3.25 [29]. The SNP data of the 72 representatives of 3,000 rice accessions corresponding to the positions of our SNP markers were obtained from the Rice SNP-Seek Database [30] and used to construct a phylogenetic tree referring to five groups of rice (*O. sativa*), i.e., *indica*, *tropical japonica*, *temperate japonica*, *aus*, and *aromatic* (Table S2; Figure S1).

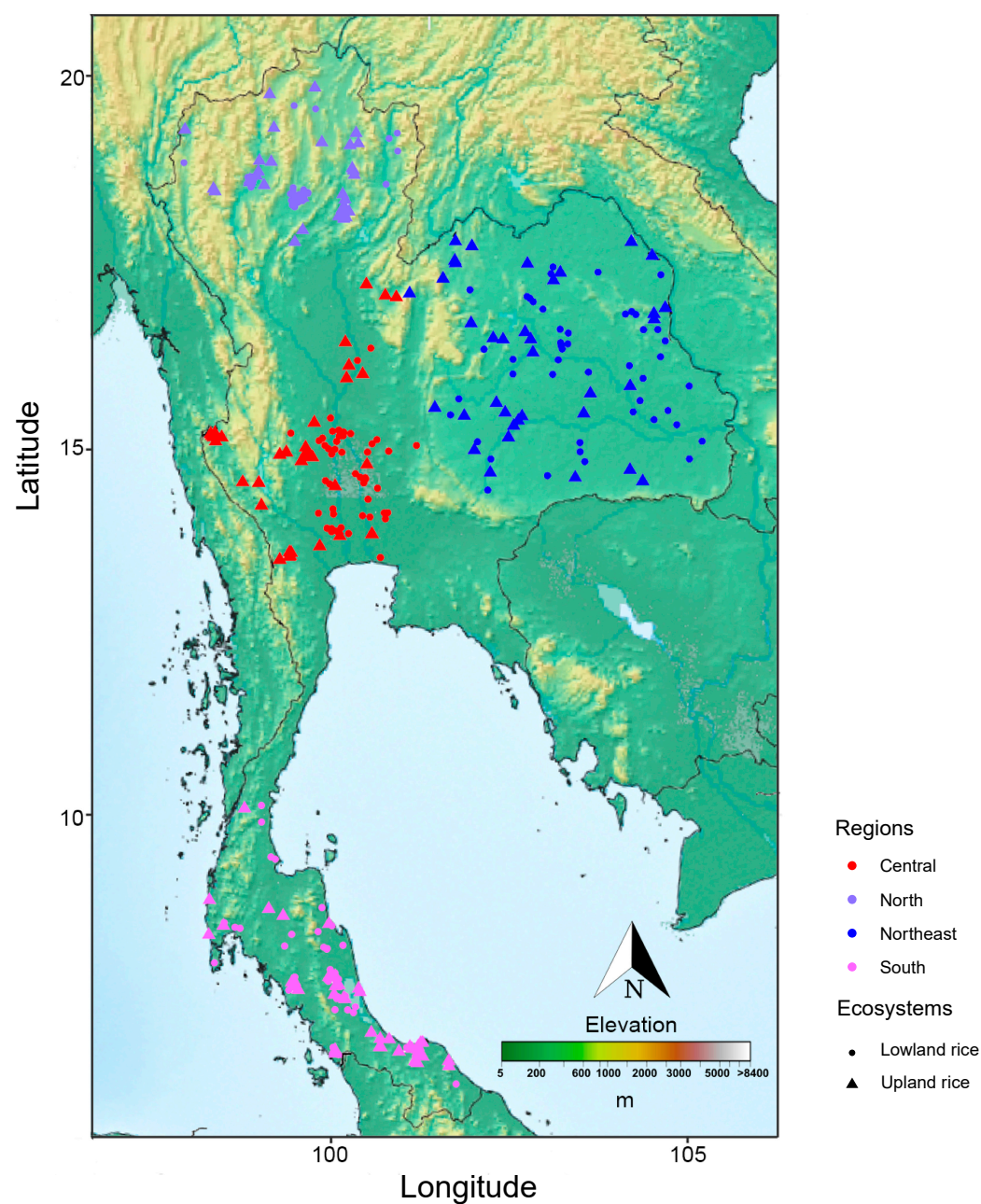
Principal Coordinate Analysis (PCoA), STRUCTURE analysis, and Discriminant Analysis of Principal Components (DAPC) were used to investigate the patterns of population structure. PCoA was performed using DARwin 6.0.021 [31], and PCs were plotted using ggplot2 [32]. The STRUCTURE analysis was performed using a Bayesian model-based clustering algorithm implemented in STRUCTURE version 2.3.4 [33], where the admixture model with the correlated allele frequencies was used. A total of 3 independent replicates were run for each genetic cluster (K) value ( $K = 1-8$ ), using a burn-in period of 100,000 and a run length of 100,000 iterations.  $\text{LnP}(D)$  values were derived for each K and plotted to find the plateau of the  $\Delta K$  [34]. The final population structure was calculated using the structure harvester [35]. DAPC was performed with the *adeigenet* package [36]. The "*find.cluster*" function was used to identify clusters (k), and the optimal k value was determined according to the Bayesian Information Criterion (BIC). Then, the "*dapc*" function was used to verify the classification quality. GenAlEx V6.5 [37] was used to compute the pairwise fixation index ( $F_{ST}$ ) among all pairs of populations, in order to investigate the population differentiation. The same tool was used to perform analysis of molecular variance (AMOVA) to estimate the components of the variances of the populations. Mantel tests were conducted using the R package *vegan* [38].

## 3. Results

### 3.1. Genetic Variability of 365 Rice Accessions Based on SNP Markers

A collection of 365 Thai rice accessions were used in this study. These included rice landraces from four geographical regions of Thailand representing different ecosystems: the upland (169 accessions) and lowland (196 accessions) (Figure 1, Table S1). To study the genetic variability among these rice accessions, a total of 102 KASP markers were initially used to genotype the 365 rice accessions (Table S3). After the removal of SNP markers that had missing genotypes above 10% or had a minor allele frequency (MAF) lower than 5%, a total of 75 SNP markers (73.5%) were retained (Table 1). These 75 SNP markers were distributed over the 12 chromosomes, in which the highest number (11 markers) and the lowest number (2 markers) of SNP markers were found on chromosomes 4 and 10, respectively (Figure 2). The polymorphic information content (PIC) ranged from 0.11 to 0.37, with an average of 0.26, and the gene diversity, also known as the expected heterozygosity ( $H_e$ ), varied from 0.12 to 0.49 (Table 1). The observed heterozygosity varied from 0 to 0.02. All SNP markers were biallelic, containing only two alleles for each marker. The

major allele frequency (MAF) at each locus ranged from 0.50 to 0.94. Most of the markers showed a variable allele frequency compared among different geographical populations (Figure S2).



**Figure 1.** The geographical distribution of the 365 rice landrace accessions across the four geographical regions of Thailand. Different colors represent different populations. Circle shapes represent lowland rice and triangle shapes represent upland rice.



**Table 1.** The polymorphism levels detected using the major allele frequency (MAF), number of different alleles ( $N_a$ ), number of effective alleles ( $N_e$ ), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), and Polymorphic Information Content (PIC) based on the SNP markers used in the study.

Marker	Chr.	Position	REF	ALT	MAF	$N_a$	$H_o$	$H_e$	PIC
R01002684973	1	2684973	C	A	0.65	2	0	0.46	0.35
R01008961781	1	8961781	C	A	0.89	2	0.01	0.19	0.17
R01023762056	1	23762056	G	A	0.7	2	0	0.42	0.33
R01033064986	1	33064986	T	C	0.91	2	0	0.16	0.15
R01033454937	1	33454937	A	T	0.64	2	0	0.46	0.36
R01034609753	1	34609753	C	G	0.93	2	0	0.13	0.12
R01036673058	1	36673058	C	T	0.91	2	0	0.17	0.15
R02008303326	2	8303326	G	A	0.89	2	0.01	0.20	0.18
R02008313120	2	8313120	G	T	0.69	2	0.01	0.43	0.34
R02008315835	2	8315835	T	A	0.94	2	0	0.12	0.11
R02010864977	2	10864977	G	T	0.82	2	0	0.30	0.25
R02019361361	2	19361361	A	C	0.9	2	0	0.18	0.17
R02019361451	2	19361451	G	T	0.9	2	0	0.18	0.17
R03016733359	3	16733359	A	C	0.72	2	0.02	0.40	0.32
R03017286744	3	17286744	G	T	0.81	2	0	0.31	0.26
R03017571575	3	17571575	G	A	0.84	2	0.01	0.27	0.23
R03031335170	3	31335170	C	T	0.84	2	0.01	0.27	0.23
R03031437173	3	31437173	C	A	0.91	2	0	0.16	0.15
R04001019335	4	1019335	A	G	0.79	2	0	0.33	0.28
R04005867320	4	5867320	T	G	0.66	2	0	0.45	0.35
R04006451939	4	6451939	C	T	0.86	2	0	0.24	0.21
R04006968050	4	6968050	A	G	0.93	2	0	0.13	0.12
R04011087541	4	11087541	G	T	0.69	2	0.02	0.43	0.34
R04011803874	4	11803874	T	A	0.58	2	0	0.49	0.37
R04016214692	4	16214692	CA	TG	0.57	2	0	0.49	0.37
R04016218749	4	16218749	T	G	0.57	2	0.01	0.49	0.37
R04022184296	4	22184296	A	G	0.64	2	0.02	0.46	0.35
R04023172729	4	23172729	T	G	0.74	2	0	0.38	0.31
R04023175725	4	23175725	T	C	0.75	2	0	0.37	0.3
R05003673333	5	3673333	T	C	0.68	2	0.01	0.44	0.34
R05019155333	5	19155333	A	G	0.59	2	0.01	0.48	0.37
R05019155705	5	19155705	G	T	0.5	2	0.01	0.50	0.37
R05023218617	5	23218617	A	C	0.7	2	0	0.42	0.33
R05026487913	5	26487913	A	G	0.91	2	0	0.17	0.15
R05028876504	5	28876504	A	T	0.91	2	0.01	0.16	0.15
R05028876779	5	28876779	A	C	0.91	2	0.01	0.16	0.15
R06001693194	6	1693194	G	A	0.54	2	0.01	0.50	0.37
R06001693411	6	1693411	A	C	0.56	2	0.01	0.49	0.37
R06001765760	6	1765760	G	T	0.54	2	0	0.50	0.37
R06001768006	6	1768006	A	C	0.79	2	0	0.33	0.27
R06001768724	6	1768724	T	C	0.82	2	0	0.29	0.25
R06001768997	6	1768997	C	T	0.69	2	0.01	0.43	0.34
R06006752886	6	6752886	G	T	0.55	2	0.01	0.49	0.37
R07005873563	7	5873563	G	A	0.79	2	0	0.33	0.28
R07020013105	7	20013105	G	T	0.55	2	0	0.49	0.37
R07020826100	7	20826100	G	T	0.82	2	0.01	0.30	0.26
R07024350575	7	24350575	C	T	0.91	2	0	0.17	0.15
R07025982551	7	25982551	A	G	0.91	2	0	0.16	0.15
R07027746661	7	27746661	C	G	0.53	2	0.01	0.50	0.37
R08002269285	8	2269285	G	C	0.56	2	0	0.49	0.37
R08002890407	8	2890407	G	A	0.63	2	0	0.46	0.36
R08003007900	8	3007900	C	G	0.54	2	0	0.50	0.37
R08020382861	8	20382861	ATTATGGC	--	0.87	2	0	0.23	0.2
R08027057202	8	27057202	G	A	0.83	2	0	0.28	0.24

Table 1. Cont.

Marker	Chr.	Position	REF	ALT	MAF	Na	Ho	He	PIC
R08027176617	8	27176617	C	A	0.72	2	0	0.40	0.32
R08027943348	8	27943348	G	A	0.86	2	0	0.24	0.21
R09004198183	9	4198183	C	A	0.77	2	0.01	0.35	0.29
R09007245205	9	7245205	A	T	0.9	2	0.01	0.18	0.16
R09007245448	9	7245448	C	T	0.9	2	0.01	0.18	0.16
R09007245650	9	7245650	A	G	0.9	2	0	0.17	0.16
R09007245739	9	7245739	C	G	0.9	2	0.01	0.17	0.16
R09007246222	9	7246223	T	G	0.9	2	0.01	0.18	0.16
R09007246804	9	7246804	T	G	0.85	2	0.01	0.26	0.23
R10015680316	10	15680316	C	A	0.92	2	0	0.15	0.14
R10021955049	10	21955049	A	G	0.9	2	0	0.18	0.16
R11013840467	11	13840467	A	G	0.81	2	0	0.30	0.26
R11014118135	11	14118135	A	G	0.6	2	0.01	0.48	0.36
R11014169508	11	14169508	C	A	0.75	2	0.01	0.37	0.3
R11021481048	11	21481048	ATT	--	0.52	2	0	0.50	0.37
R11028209211	11	28209211	A	T	0.7	2	0	0.42	0.33
R12009613648	12	9613648	A	C	0.9	2	0.01	0.18	0.17
R12018759236	12	18759236	T	G	0.85	2	0	0.25	0.22
R12018856177	12	18856177	G	A	0.83	2	0	0.29	0.24
R12022373643	12	22373643	A	G	0.69	2	0.01	0.43	0.34
R12023173265	12	23173265	T	A	0.77	2	0	0.36	0.29

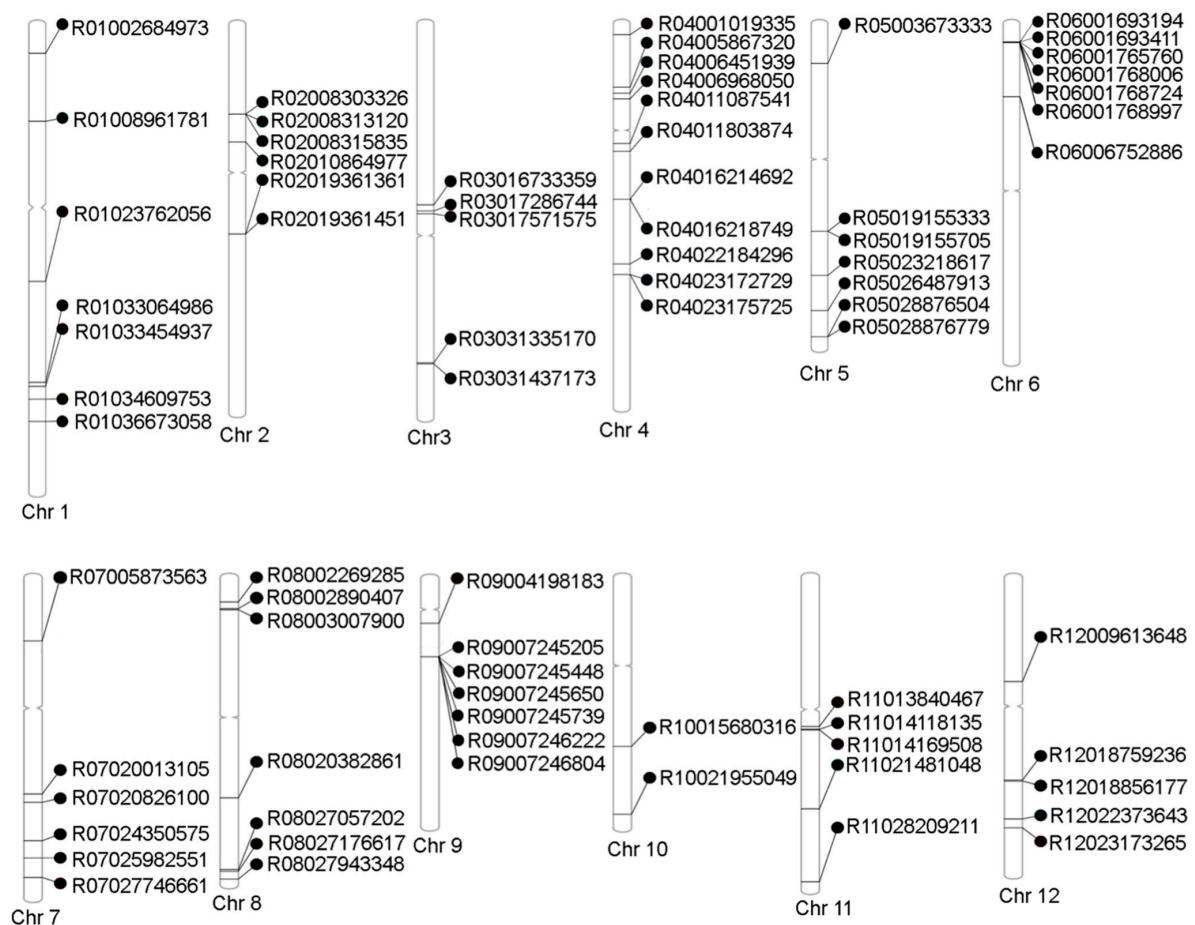
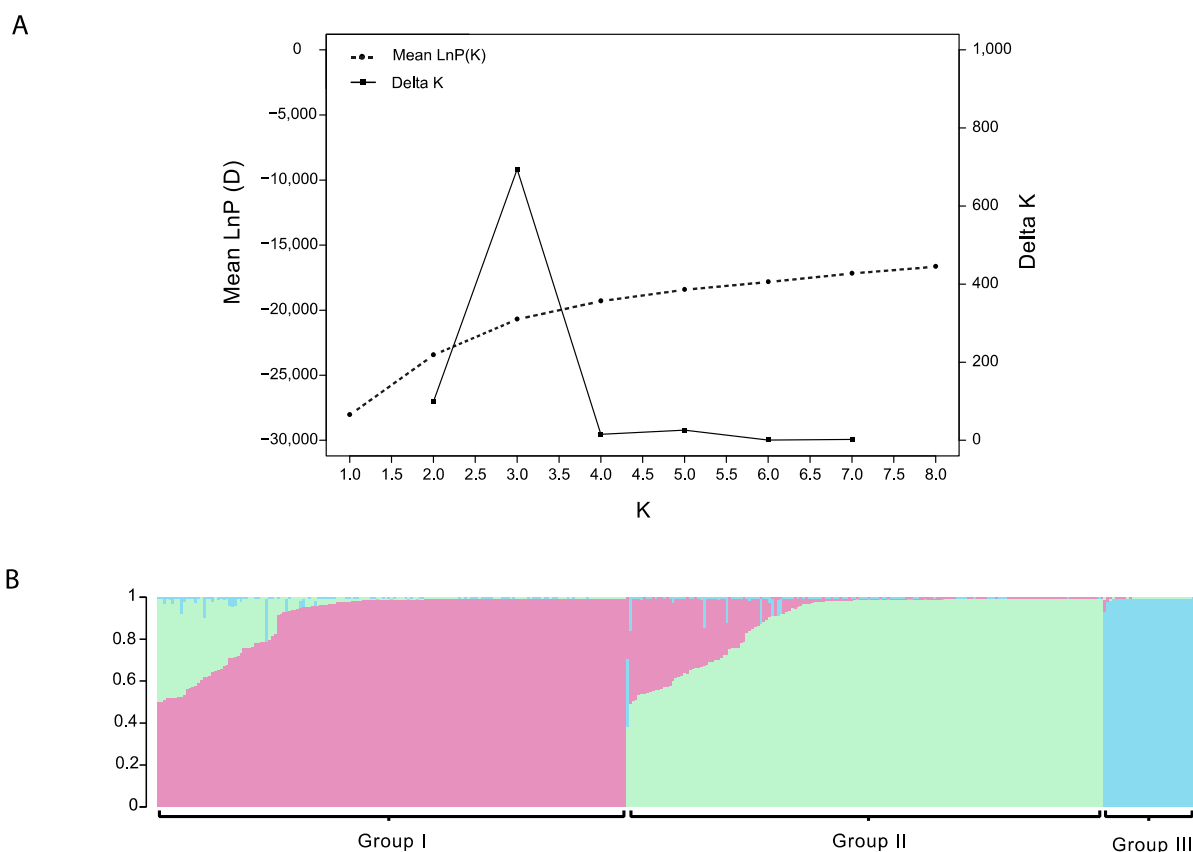


Figure 2. Distribution of the SNP markers used in this study across the 12 rice chromosomes.

### 3.2. Population Structure of Thai Landraces

To understand the structure of the overall 365 rice accessions among the four different geographical regions, three different approaches, i.e., STRUCTURE, DAPC, and PCoA, were applied. All analyses were performed based on the genotypes identified by the 75 SNP markers. In STRUCTURE, the number of clusters was estimated based on the  $\Delta K$  method [34] and the plateau criterion [33]. The results suggested the best grouping number at  $K = 3$  based on the  $\Delta K$  and the mean log-likelihood ( $\text{LnP}(D)$ ) curve (Figure 3). This suggests that the 365 rice landraces can be grouped into three subpopulations, referred to as Groups I–III (Figure 3).



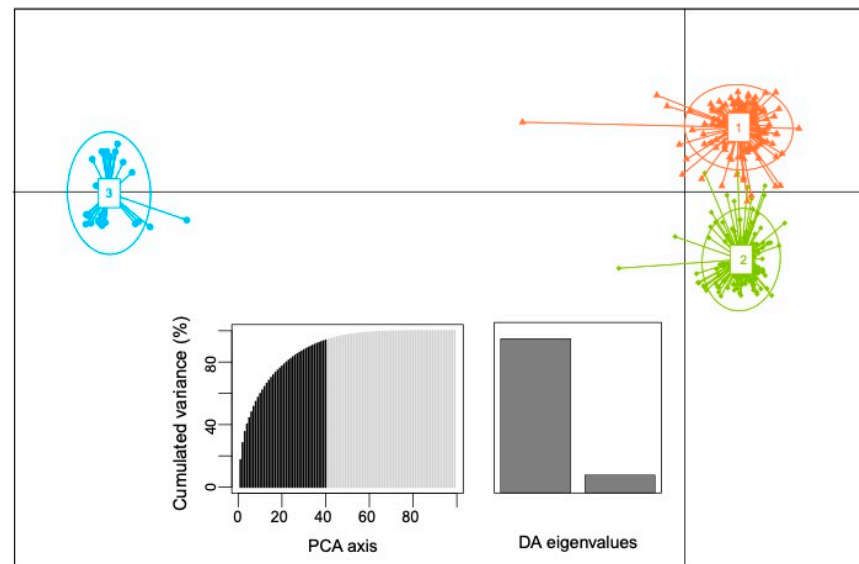
**Figure 3.** The population structure of 365 rice accessions at  $K = 3$ . (A) The average log-likelihood of the  $K$ -value against the number of  $K$  and delta  $K$  for different numbers of subpopulations ( $K$ ); (B) the estimated population structure of the 365 Thai landrace accessions on  $K = 3$ .

In each group, accessions with a score higher than 0.80 were assigned to a pure group, while those with a score lower than 0.80 were assigned to be admixture. We found that the accessions assigned into Groups I and II were all *indica* landraces (Table S4). Group I consisted of 168 accessions where 45.23% were landraces from the southern (S) population and 33.33% were landraces from the central (C) population. Within this cluster, 126 accessions contained pure genotypes, and 42 accessions exhibited an admixed ancestry.

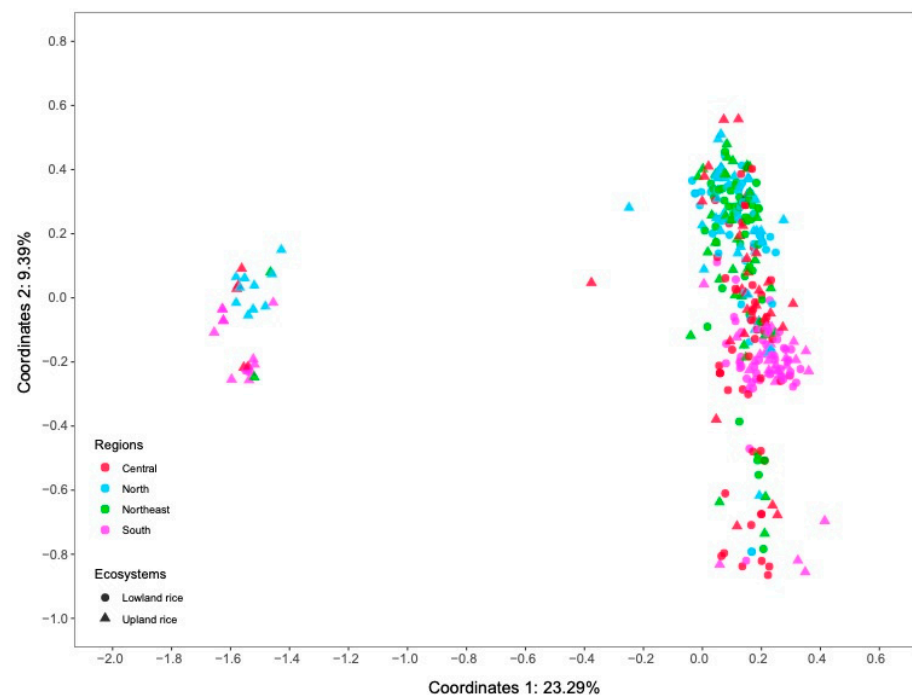
Group II consisted of 165 accessions, 81.21% of which were landraces from the northern (N) and the northeastern (NE) populations. Within this group, 126 accessions contained pure genotypes, and 39 accessions exhibited an admixed ancestry. Group III contained 32 accessions, all of which were *japonica* landraces, and most of which were upland rice from the N and S populations (Table S4). None of the accessions in this group contained admixed ancestry.

The number of subpopulations and the subpopulation attributions yielded by DAPC were similar to those revealed with STRUCTURE (Figure 4). The three groups were clearly

defined as two closely related indica subgroups (Groups 1 and 2) and a distantly related japonica group (Group 3) (Figure 4; Table S4). Similarly, the PCoA result revealed two main clusters corresponding to the two groups, indica and japonica. However, subgroups among the indica landraces were not clearly defined (Figure 5; Table S4).



**Figure 4.** A discriminant analysis of principle components (DAPC) scatter plot of the 365 accessions of Thai rice landraces. The x and y axes represent the discriminant functions one and two, respectively. Dots represent the individual samples, and centroids are indicated with numbers (1–3). The inset bar charts indicate the number of principle components retained by DAPC (right) and the cumulative variation (left).

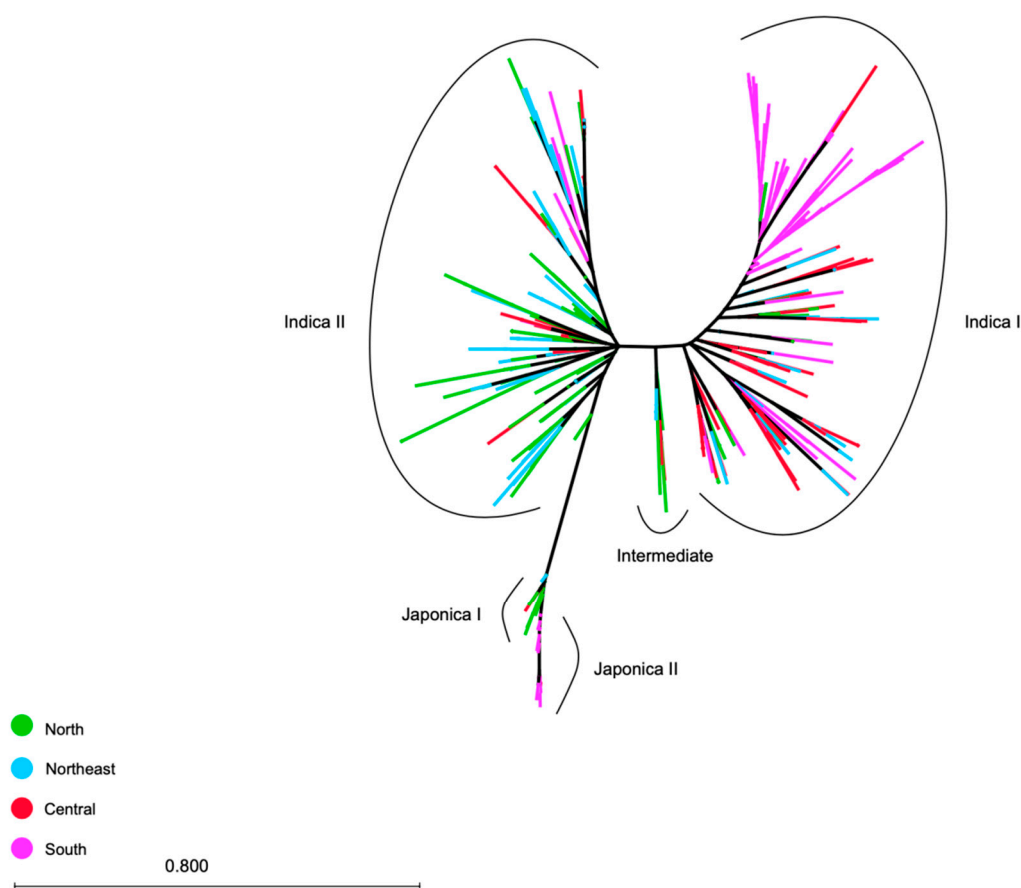


**Figure 5.** Principal coordinate analysis (PCoA) ordination based on the Nie distance matrix of the genotypes from 75 SNP markers comparing 365 samples. The PCoA is colored according to the regional populations. The circle and triangle symbols depict lowland and upland ecosystems, respectively. The percentage of the variation coordinate is indicated on the axes.



### 3.3. Genetic Distance and Phylogeny of the 365 Accessions

To examine the genetic relationships among the 365 rice landraces, we calculated the genetic distance based on Nie's dissimilarity. As a result, two major clusters of these rice landraces, corresponding to *indica* and *japonica* groups, were revealed (Figure S3). The phylogenetic tree constructed using the maximum likelihood method clustered these landraces into two major groups, corresponding to *indica* and *japonica*, similar to the groups identified by DAPC and PCoA (Figure 6). Among the *indica* group, two major clusters (namely Indica I and II) were further observed. The Indica I subgroups contained 176 accessions, 74 and 61 of which were rice landraces from the S population and the C population, respectively. Most of the rice landraces in this group were non-glutinous rice (Table S4). The *indica* II subgroups contained 148 accessions, 122 of which were rice landraces from the N and NE populations. The majority of the rice landraces in this group (72.29%) were glutinous rice. The landraces from the S population were rarely contained in this group (Table S4). There was also a small intermediate group present between the two *indica* subgroups. This intermediate *indica* group contained nine rice landraces from the C, N, and NE populations (Table S4). Among the *japonica* group, two clusters (namely Japonica I and II) were indicated. The Japonica I subgroup predominantly contained the rice accessions from the N population (10/14 accessions), while the Japonica II subgroup predominantly contained the accessions from the S population (15/18 accessions).



**Figure 6.** Maximum-likelihood dendrogram illustrating the genetic relationship among the 365 landrace accessions using four-region population partitioning.

While both upland and lowland rice landraces were found in all *indica* subgroups, the upland rice landraces were predominantly included in the *japonica* group (Table S4). The overall genetic diversity of the 365 rice accessions was moderate, as revealed by the mean gene diversity (0.33) and mean PIC (0.26) (Table 2). Similar values of gene diversity

(0.28–0.32) and PIC (0.23–0.25) were found among the four geographical populations (Table 2). However, the mean genetic diversity of the japonica group (0.09) was much lower than that of the other two indica subgroups (0.27 and 0.24 for Indica I and II, respectively) (Table 3).

**Table 2.** The genetic diversity parameters of four subpopulations grouped by the four regions of Thailand.

Population	Sample Size	Major Allele Frequency/Locus	Mean Gene Diversity	Mean PIC Value
All subpopulations	365	0.76	0.33	0.26
Central (C)	89	0.77	0.32	0.25
Northern (N)	96	0.80	0.29	0.24
Northeastern (NE)	86	0.81	0.28	0.23
Southern (S)	94	0.77	0.31	0.25

**Table 3.** Genetic diversity parameters of the three subpopulations grouped by STRUCTURE analysis.

Population	Sample Size	Major Allele Frequency /Locus	Mean Gene Diversity	Mean PIC Value
All subpopulations	365	0.76	0.32	0.26
Group I (indica)	167	0.80	0.27	0.21
Group II (indica)	166	0.82	0.24	0.19
Group III (japonica)	32	0.93	0.09	0.07

### 3.4. Genetic Differentiation, AMOVA, and Isolation-by-Distance Analyses

We further quantified the genetic differentiation ( $F_{ST}$ ) between each geographical population using Wright's test [39]. In practice, an  $F_{ST}$  of 0.00–0.05 indicates low differentiation, 0.05–0.15 indicates moderate differentiation, and  $F_{ST}$  of >0.15 indicates a high level of differentiation [40]. Pairwise estimates of the  $F_{ST}$  values between pairs of subpopulations due to geographical regions revealed low to moderate genetic differentiation, ranging from 0.016 to 0.078 (Table 4). The largest genetic differentiation ( $F_{ST}$ ) was detected between the NE and S populations (0.078), and the lowest genetic differentiation was detected between the N and NE populations (0.016). The levels of genetic differentiation between the S and N populations and between the S and C populations were also moderate as indicated by the  $F_{ST}$  values of 0.069 and 0.052, respectively.

**Table 4.** The pairwise population  $F_{ST}$  analysis of the four subpopulations grouped by region of Thailand.

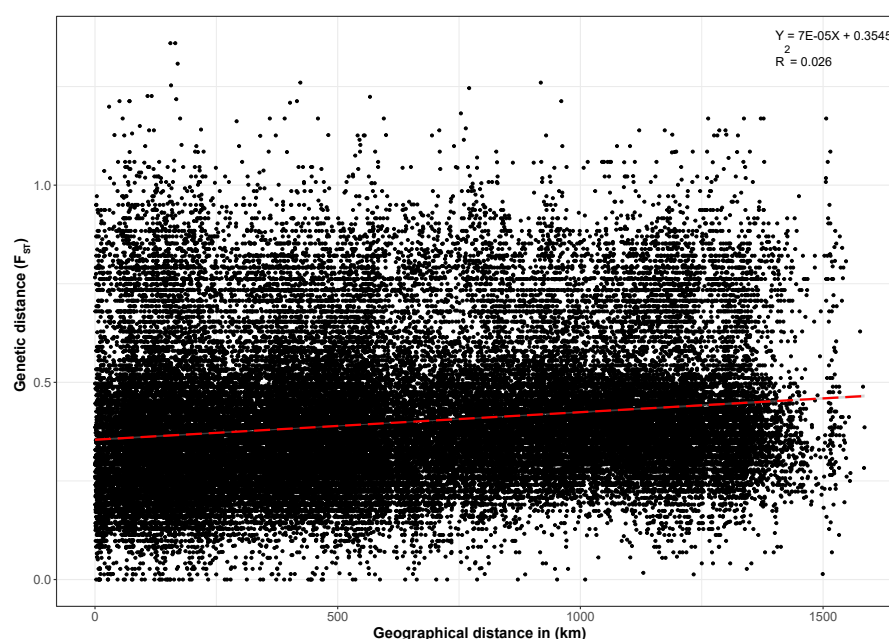
Subpopulations	Central	North	Northeast	South
Central (C)	0.000			
Northern (N)	0.039	0.000		
Northeast (NE)	0.023	0.016	0.000	
South (S)	0.052	0.069	0.078	0.000

The analysis of molecular variance (AMOVA) based on the geographical regions revealed significant variations ( $p < 0.001$ ) among and within the subpopulation groups corresponding to geographic regions (Table 5). Differences within the geographical regions contributed to approximately 91% of the total genetic variations, and these differences were notably and significantly higher than those among regions (only 9% of the total genetic variation was due to differences among regions; Table 5).

**Table 5.** The AMOVA analysis for the geographical populations in the collection of 365 landraces.

Source	df	SS	MS	Estimated Variance	Percentage of Total Variance	Probability ( <i>p</i> )
Among Pops	3	1463.724	487.908	4.835	9%	<0.001
Within Pops	361	16970.874	47.011	47.011	91%	
Total	364	18434.597		51.846	100%	

Isolation-by-distance (IBD) based on Mantel's test was also demonstrated when comparing the genetic distances ( $F_{ST}$ ) and geographical distances. A significant positive relationship between the geographical and genetic distances ( $R^2 = 0.026$ ,  $p < 0.001$ ) was observed in the entire population based on the SNP data (Figure 7). A similar relationship was also observed between the genetic distances ( $F_{ST}$ ) and some climatic factors, i.e., latitude, elevation, temperature, and rainfall, among the 365 accessions (Figures S4–S7).

**Figure 7.** Correlation of the geographical distance (in kilometers) and genetic distance (pairwise  $F_{ST}$ ) among the 365 rice accessions.

#### 4. Discussion

This study provides a comprehensive and systematic study of the genetic variation in a large collection of Thai rice landraces. The panel consisted of 365 accessions, which included 169 and 196 accessions of upland and lowland rice, respectively. These accessions were representatives of landraces obtained from four geographic origins: northern (96 accessions from 9 provinces), northeastern (86 accessions from 17 provinces), central (89 accessions from 15 provinces), and southern Thailand (94 accessions from 14 provinces). Based on the 75 SNP markers, the PIC values for all SNPs were less than 0.5, with an average PIC value of 0.26, suggesting that all the SNPs were considered moderate- or low-information markers. This may be due to the bi-allelic nature of the SNPs [41].

The different approaches (STRUCTURE, PCoA, and DAPC) used to analyze the genetic structure of the 365 rice accessions appeared to provide complementary information. The results showed a good consistency between STRUCTURE and DAPC. The population structure analysis using these approaches suggested that the Thai rice landraces could be divided into two main subpopulations, corresponding to *indica* and *japonica* groups, similar to the previous reports [21,23,24]. However, the majority of accessions present in our panel were *indica* rice. The genetic diversity was higher for the *indica* groups compared to the *japonica* group.

While both the upland and lowland landraces were distributed among the *indica* groups, most landraces in the *japonica* group were upland rice. In Thailand, *indica* is the dominant rice type and is normally grown in the lowland ecosystem. On the other hand, *japonica* rice is rarely found in mainstream rice production in Thailand and is, instead, mostly enriched in regions, such as the northern highlands [42].

Based on the rice collection used in this study, *japonica* rice was likely to be separated into two subgroups, one specific to the northern population and another specific to the southern population. It is possible that both subgroups of Thai *japonica* rice arose in both regions from different origins. There were also some rice accessions from the central population belonging to the Japonica I subgroup (Figure 6; Table S4). These rice accessions may have the same origin as those in the northern population and may be exchanged by farmers between the two regions. It is also worth determining whether the two classes of *japonica* have different physiological properties.

$F_{ST}$  measures the amount of genetic variance that can be explained by the population structure based on Wright's F-statistics [39]. An  $F_{ST}$  value of 0 indicates no differentiation between the subpopulations, while a value of 1 indicates complete differentiation. An  $F_{ST}$  value greater than 0.15 can be considered as significant in differentiating populations [43]. In the present study, no significant divergence was found among the four regional populations (Table 2). This coincided with the AMOVA results (Table 5), where the vast majority of the total variation (91%) was accounted for by within-subpopulation variation, while only 9% of the total variation was accounted for by among-subpopulation variation. Assuming germplasm exchanges among regions, an independent clustering pattern of landraces from different origins in the maximum-likelihood phylogenetic tree analysis, and the revealed high molecular variance among accessions within populations in AMOVA pointed to human pressure on the dissemination of seeds among populations. This is also supported by the isolation-by-distance analysis result where only a low level of correlation was detected between the genetic distance ( $F_{ST}$ ) and geographic distance for the Thai rice landraces.

## 5. Conclusions

In this study, the genetic diversity and population structure of 365 Thai rice landraces were investigated using a set of 75 SNP markers. Two main subpopulations, corresponding to *indica* and *japonica*, were clearly identified across the whole panel of these rice accessions. Among the *indica* group, two further subgroups were also indicated. The genetic differentiation among the four geographic populations was not high, which may be due to the occurrence of genetic exchange between populations.

This study provides a detailed understanding of the genetic structure and diversity of Thai landraces, which is crucial for the efficient utilization of rice genetic resources and for developing suitable conservation strategies. This will aid in building a comprehensive collection of landraces in terms of genetic diversity, which is fundamental for other studies, such as genome-wide association studies (GWAS).

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/agronomy11050995/s1>, Figure S1: The maximum-likelihood tree of the 72 representative of 3000 rice genomes (highlighted in different colors) and the 365 Thai landrace accessions, Figure S2: Heatmap hierarchical clustering of the genetic distance calculated based on the dissimilarity matrix of 365 rice accessions, Figure S3: Allele frequency of the 75 SNP markers in each geographical population, Figure S4: Correlation of the latitude distance and genetic distance (pairwise  $F_{ST}$ ) among the 365 rice accessions, Figure S5: Correlation of the pairwise elevation difference (in meters) and genetic distance (pairwise  $F_{ST}$ ) among the 365 rice accessions, Figure S6: Correlation of the pairwise difference of average temperature in a year (in degree Celsius) and genetic distance (pairwise  $F_{ST}$ ) among the 365 rice accessions, Figure S7: Correlation of the pairwise difference of average rainfall in a year (in millimeters) and genetic distance (pairwise  $F_{ST}$ ) among the 365 rice accessions, Table S1: List of rice accessions used in the study, Table S2: List of 72 representative of 3000 rice genomes assigned to five genetic groups, Table S3: List of 102 SNP markers used for genotyping, and Table S4: List of 365 rice accessions used in the study.



**Author Contributions:** Conceptualization, T.S., C.M., A.V., T.T., S.W., and S.A.; Formal analysis, W.A., S.R., and V.R.; Funding acquisition, T.T., T.S., and S.A.; Methodology, W.A. and S.W.; Supervision, T.T., S.W., and S.A.; Writing—original draft, W.A. and S.W.; Writing—review & editing, S.W. and S.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by the Center of Excellence on Agricultural Biotechnology, Science and Technology Postgraduate Education and Research Development Office, Office of Higher Education Commission, Ministry of Higher Education, Science, Research and Innovation. (AG-BIO/PERDO-CHE), the National Research Council of Thailand (NRCT), grant numbers NRCT5-RSA63002-06 and NRCT-RTA/812/2563, and the Kasetsart University Research and Development Institute (KURDI), Kasetsart University.

**Data Availability Statement:** The authors confirm that the data supporting the findings of this study are available within the article and its Supplementary Materials.

**Acknowledgments:** The authors would also like to thank the Thailand Rice Department, Ministry of Agriculture and Cooperatives, for rice material support.

**Conflicts of Interest:** The authors declare that they have no conflict of interests.

## References

1. Khush, G.S. Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **1997**, *35*, 25–34. [\[CrossRef\]](#)
2. Bandumula, N. Rice production in asia: Key to global food security. *Proc. Natl. Acad. Sci. USA* **2018**, *88*, 1323–1328. [\[CrossRef\]](#)
3. Silva, C.N. *Urban Planning in Sub-Saharan Africa: Colonial and Post-Colonial Planning Cultures*; Routledge: London, UK, 2015; ISBN 9781315797311.
4. Oka, H.-I.; Morishima, H. Phylogenetic differentiation of cultivated rice, XXIII. Potentiality of wild progenitors to evolve the Indica and Japonica types of rice cultivars. *Euphytica* **1982**, *31*, 41–50. [\[CrossRef\]](#)
5. Mackill, D.J.; Lei, X. Genetic variation for traits related to temperate adaptation of rice cultivars. *Crop Sci.* **1997**, *37*, 1340–1346. [\[CrossRef\]](#)
6. Mackill, D.J.; Khush, G.S. IR64: A high-quality and high-yielding mega variety. *Rice* **2018**, *11*, 18. [\[CrossRef\]](#)
7. Singh, R.K. Genetic resource and the role of international collaboration in rice breeding. *Genome* **1999**, *42*, 635–641. [\[CrossRef\]](#)
8. Harlan, J.R. Our vanishing genetic resources. *Science* **1975**, *188*, 617–621. [\[CrossRef\]](#)
9. Kovach, M.J.; McCouch, S.R. Leveraging natural diversity: Back through the bottleneck. *Curr. Opin. Plant Biol.* **2008**, *11*, 193–200. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Govindaraj, M.; Vetriventhan, M.; Srinivasan, M. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. *Genet Res Int* **2015**, *2015*, 431487. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Bhandari, H.R.; Bhanu, A.N.; Srivastava, K.; Singh, M.N. Assessment of genetic diversity in crop plants-an overview. *Adv. Plants Agric.* **2017**, *7*, 255.
12. Glaszmann, J.C. Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* **1987**, *74*, 21–30. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Garriss, A.J.; Tai, T.H.; Coburn, J.; Kresovich, S.; McCouch, S. Genetic structure and diversity in *Oryza sativa* L. *Genetics* **2005**, *169*, 1631–1638. [\[CrossRef\]](#)
14. Sow, M.; Ndjioudjop, M.-N.; Sido, A.; Mariac, C.; Laing, M.; Bezançon, G. Genetic diversity, population structure and differentiation of rice species from Niger and their potential for rice genetic resources conservation and enhancement. *Genet. Resour. Crop Evol.* **2014**, *61*, 199–213. [\[CrossRef\]](#)
15. Caicedo, A.L.; Williamson, S.H.; Hernandez, R.D.; Boyko, A.; Fledel-Alon, A.; York, T.L.; Polato, N.R.; Olsen, K.M.; Nielsen, R.; McCouch, S.R.; et al. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet.* **2007**, *3*, 1745–1756. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Huang, X.; Kurata, N.; Wei, X.; Wang, Z.-X.; Wang, A.; Zhao, Q.; Zhao, Y.; Liu, K.; Lu, H.; Li, W.; et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **2012**, *490*, 497–501. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Tam, N.T.; Dwiyantri, M.S.; Koide, Y.; Nagano, A.J.; Ky, H.; Tin, H.Q.; Hien, N.L.; Dung, L.V.; Kishima, Y. Profiling SNP and nucleotide diversity to characterize mekong delta rice landraces in southeast asian populations. *Plant Genome* **2019**, *12*, 190042. [\[CrossRef\]](#)
18. Zhao, K.; Wright, M.; Kimball, J.; Eizenga, G.; McClung, A.; Kovach, M.; Tyagi, W.; Ali, M.L.; Tung, C.-W.; Reynolds, A.; et al. Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS ONE* **2010**, *5*, e10780. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Sang, T.; Ge, S. Genetics and phylogenetics of rice domestication. *Curr. Opin. Genet. Dev.* **2007**, *17*, 533–538. [\[CrossRef\]](#)
20. Sweeney, M.; McCouch, S. The complex history of the domestication of rice. *Ann. Bot.* **2007**, *100*, 951–957. [\[CrossRef\]](#)
21. Kladmook, M.; Kumchoo, T.; Hongtrakul, V. Genetic diversity analysis and subspecies classification of Thailand rice landraces using DNA markers. *Afr. J. Biotechnol.* **2012**, *11*, 14044–14053. [\[CrossRef\]](#)
22. Nilthong, S.; Chukeatirote, E. Assessment of genetic diversity in Thai upland rice varieties using SSR markers. *Aust. J. Crop. Sci.* **2020**. [\[CrossRef\]](#)

23. Pathaichindachote, W.; Panyawut, N.; Sikaewtung, K.; Patarapuwadol, S.; Muangprom, A. Genetic diversity and allelic frequency of selected thai and exotic rice germplasm using SSR markers. *Rice Sci.* **2019**, *26*, 393–403. [\[CrossRef\]](#)
24. Chakhonkaen, S.; Pitnjam, K.; Saisuk, W.; Ukoskit, K.; Muangprom, A. Genetic structure of Thai rice and rice accessions obtained from the International Rice Research Institute. *Rice* **2012**, *5*, 19. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Pusadee, T.; Wongtamee, A.; Rerkasem, B.; Olsen, K.M.; Jamjod, S. Farmers Drive Genetic Diversity of Thai Purple Rice (*Oryza sativa* L.) Landraces. *Econ Bot* **2019**, *73*, 76–85. [\[CrossRef\]](#)
26. Nei, M. Genetic Distance between Populations. *Am. Nat.* **1972**, *106*, 283–292. [\[CrossRef\]](#)
27. Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **1993**, *10*, 512–526. [\[CrossRef\]](#)
28. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [\[CrossRef\]](#)
29. Liu, K.; Muse, S.V. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [\[CrossRef\]](#)
30. Alexandrov, N.; Tai, S.; Wang, W.; Mansueto, L.; Palis, K.; Fuentes, R.R.; Ulat, V.J.; Chebotarov, D.; Zhang, G.; Li, Z.; et al. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* **2015**, *43*, D1023–D1027. [\[CrossRef\]](#)
31. Perrier, X.; Jacquemoud-Collet, J.P. DARwin software. 2006. Available online: <https://darwin.cirad.fr> (accessed on 4 May 2021).
32. Villanueva, R.A.M.; Chen, Z.J. ggplot2: Elegant Graphics for Data Analysis (2nd ed.). *Meas. Interdiscip. Res. Perspect.* **2019**, *17*, 160–167. [\[CrossRef\]](#)
33. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [\[CrossRef\]](#)
34. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [\[CrossRef\]](#)
35. Earl, D.A.; von Holdt, B.M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [\[CrossRef\]](#)
36. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403–1405. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Peakall, R.; Genalex, S. 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Bioinformatics* **2012**, *28*, 19. [\[CrossRef\]](#)
38. Oksanen, J.; Blanchet, F.G.; Kindt, R.; Legendre, P. Vegan: Community Ecology Package. R-Package Version 2.0-10. 2013. Available online: <http://CRAN.R> (accessed on 4 May 2021).
39. Wright, S. Evolution and the genetics of populations, volume 4: Variability within and among natural populations. *Biometrics* **1979**, *35*, 359. [\[CrossRef\]](#)
40. Govindaraju, D.R. Variation in gene flow levels among predominantly self-pollinated plants. *J. Evol. Biol.* **1989**, *2*, 173–181. [\[CrossRef\]](#)
41. Eltaher, S.; Sallam, A.; Belamkar, V.; Emara, H.A.; Nower, A.A.; Salem, K.F.M.; Poland, J.; Baenziger, P.S. Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. *Front. Genet.* **2018**, *9*, 76. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Oka, H.I.; Chang, W.T. A note on rice varieties of Japonica type found in northern Thailand. *Bot. Bull. Acad. Sinica* **1963**, *4*, 163–168.
43. Frankham, R.; Ballou, J.D.; Briscoe, D.A.; McInnes, K.H. *Introduction to Conservation Genetics*; Cambridge University Press: Cambridge, UK, 2002; ISBN 9780511808999.