



## Article

# Genetic Diversity and Population Structure of a Longan Germplasm in Thailand Revealed by Genotyping-By-Sequencing (GBS)

Kanamon Riangwong<sup>1</sup>, Chatree Saensuk<sup>2</sup>, Mutiara K. Pitaloka<sup>2</sup>, Reajina Dumhai<sup>2</sup>, Vinitchan Ruanjaichon<sup>3</sup> , Theerayut Toojinda<sup>3</sup>, Samart Wanchana<sup>3</sup> and Siwaret Arikrit<sup>2,4,\*</sup>

<sup>1</sup> Department of Biotechnology, Faculty of Engineering and Industrial Technology, Silpakorn University, Sanamchandra Palace Campus, Nakhon Pathom 73000, Thailand; kanamonnueng@gmail.com

<sup>2</sup> Rice Science Center, Kasetsart University Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand; chatsaen@gmail.com (C.S.); mutiarakp@gmail.com (M.K.P.); reajina.d06@gmail.com (R.D.)

<sup>3</sup> National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani 12120, Thailand; vinitchan.rua@biotec.or.th (V.R.); theerayut@biotec.or.th (T.T.); samart.wan@biotec.or.th (S.W.)

<sup>4</sup> Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand

\* Correspondence: siwaret.a@ku.th

**Abstract:** Longan (*Dimocarpus longan* Lour.) is grown commercially in many countries, including China, Thailand, the Philippines, Malaysia, Vietnam, India, Australia, and Hawaii. Thailand is the second largest producer and largest exporter of longan in the world. Currently, there is limited information on the genetic background, population structure, and genetic relationships among longan cultivars in Thailand. In this study, a total of 50 longan accessions from a community-based germplasm collection in Thailand were analyzed using 10,619 SNPs from genotyping-by-sequencing (GBS). Based on the results of STRUCTURE analysis, 43 accessions were classified into 4 subpopulations, and the other 7 accessions were found to contain admixed genotypes. Based on UPGMA clustering analysis and PCoA analysis, the longan accessions could be divided into six major groups consistent with those identified by STRUCTURE. A relatively high degree of genetic variation was observed among the longan accessions, as quantified by the expected heterozygosity ( $H_e = 0.308$ ). AMOVA results showed that 74% and 26% of the total variation occurred between and within populations, respectively. Obvious genetic differentiation between populations ( $F_{ST} = 0.25$ ) was observed. The results of this study are useful for managing longan germplasm and may facilitate the genetic improvement of longan.

**Keywords:** genetic diversity; population structure; genotyping-by-sequencing; single-nucleotide polymorphism; longan



**Citation:** Riangwong, K.; Saensuk, C.; Pitaloka, M.K.; Dumhai, R.; Ruanjaichon, V.; Toojinda, T.; Wanchana, S.; Arikrit, S. Genetic Diversity and Population Structure of a Longan Germplasm in Thailand Revealed by Genotyping-By-Sequencing (GBS). *Horticulturae* **2023**, *9*, 726. <https://doi.org/10.3390/horticulturae9060726>

Received: 2 June 2023

Revised: 16 June 2023

Accepted: 19 June 2023

Published: 20 June 2023



**Copyright:** © 2023 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/>).

## 1. Introduction

Longan (*Dimocarpus longan* Lour.) is a subtropical evergreen tree belonging to the Sapindaceae family [1]. It has a diploid genome ( $2n = 2x = 30$ ) with a size of 470 Mb [2]. Longan fruit can be consumed as fresh pulp or as processed pulp products such as dried longan pulp, dried whole longan with peel, canned longan in syrup, longan jelly, longan wine, and longan juice [3,4]. Longan is commercially cultivated in many countries, including China, Thailand, the Philippines, Malaysia, Vietnam, India, Australia, and Hawaii [5]. China is the largest producer of longan, followed by Thailand. Thailand is currently the largest exporter of longan in the world. In 2021, Thailand exported about 700,000 tons of longan, worth more than \$800 million (Office of Agricultural Economics, 2021: <https://www.oae.go.th>; accessed date: 1 June 2023). Longan was introduced from China into Thailand in the late 1800s and is now considered an important economic crop with high export value to the

country [6]. In Thailand, the main cultivation areas for longan are in the north, but it is also grown in the east and northeast [7]. Many longan cultivars are grown in Thailand, including landraces and improved cultivars, as well as some foreign cultivars. The main cultivars of longan in Thailand include E-daw or Daw, Chompoo or Si Chompoo, Biew Khiew, and Haew [6]. E-daw is the most popular longan cultivar in Thailand, grown mainly in the main areas in the north and east, accounting for 90% of commercially grown longan cultivars [8].

Knowledge of the genetic background and genetic structure of populations is useful for germplasm conservation and breeding longans. Traditionally, longan cultivars are characterized by morphological traits, such as fruit characteristics [9]. Little information on genetic background is available for most commercially grown longan cultivars in Thailand. In addition, the genetic relationship among longan cultivars within Thailand has not been well characterized. Molecular markers are commonly used to assess the genetic diversity of germplasm because they are not affected by environmental factors [10]. The use of molecular markers to accurately identify longan cultivars is useful for longan germplasm management and breeding. Molecular markers that have been used to assess longan genetic diversity include amplified fragment length polymorphism (AFLP) [11], random amplified polymorphic DNA (RAPD) [12], sequence-characterized amplified region [13], sequence-related amplified polymorphism (SRAP) [14], microsatellites [15], and single-nucleotide polymorphisms (SNPs) [16]. As the most abundant type of sequence variation, SNPs are suitable for a variety of applications, including analyses of genetic diversity and population structure [17]. With advances in next-generation sequencing technology, SNP marker discovery through whole-genome resequencing or genotyping-by-sequencing is becoming easier. GBS has been widely used in plant breeding programs because it is a cost-effective and efficient high-throughput strategy for SNP discovery and genotyping. GBS can be performed with or without a genome reference [18]. It has been used for genotyping in various crops such as rice [19], maize [20], barley [21], wheat [22], potato [23], and others. In addition to SNPs, SSRs can also be inferred from GBS sequencing data [24].

In this study, we applied GBS analysis to characterize 50 longan accessions from the Nong Chang Khuen Longan Community BioBank, Lamphun, the first longan diversity biobank in Thailand. We assessed the genetic diversity and population structure of these longan accessions using a total of 10,619 GBS-derived SNPs. Our results provide insight into longan diversity and population structure, which will be useful for longan germplasm management and facilitate longan genetic improvement.

## 2. Materials and Methods

### 2.1. Plant Materials and DNA Preparation

A total of 50 longan accessions from the Longan Community BioBank, Nong Chang Khuen, Lamphun, Thailand, were used in this study (Table S1). Of these, 48 accessions were Thai cultivars, and the other 2 cultivars, Chuliang and Kohala, were foreign cultivars from China and the USA, respectively. The genomic DNA of all 50 longan accessions was isolated from 100 mg of young leaf tissue using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Then, the isolated DNA of each sample was quantified using a Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific, MA, USA) and adjusted to a final concentration of 50 ng/ $\mu$ L.

### 2.2. Genotyping-By-Sequencing (GBS) and SNP Filtering

The genomic DNA (50  $\mu$ L) was digested with the type II restriction endonuclease *ApeKI* to construct GBS libraries according to the protocol described in [21]. The library was sequenced on an Illumina HiSeq 2000 platform at Novogene (HK) Co., Ltd. (Tsim Sha Tsui, Hong Kong). Sequencing data were processed by removing low-quality sequences using Trimmomatic (v. 0.39) [25]. The cleaned reads were then aligned to the draft genome of longan [26] using Bowtie2 (v. 2.3.5.1) [27] with default parameters. Only unique matches to the genome were maintained for further analyses. SNPs were called using the Genome

Analysis Toolkit (GATK, version 3.8.1) components: HaplotypeCaller, CombineGVCFs, and GenotypeGVCFs [28]. SNPs without missing calls and MAF > 2% were used for subsequent analyses.

### 2.3. Marker Diversity Analysis

The genetic data were analyzed using multiple software combinations. The expected heterozygosity (He), polymorphism information content (PIC), minor allele frequency (MAF), and observed heterozygosity (Ho) were calculated using adegenet packages [29] in R software (<https://www.r-project.org>; accessed date: 1 June 2023).

### 2.4. Population Structure Analysis

The genetic distances of the 50 longan genotypes were calculated using Nei's standard dissimilarity distance based on 10,619 filtered SNPs, and an unweighted pair group method UPGMA phylogenetic tree was performed with the Tamura–Nei model and 500 bootstraps using MEGA X [30]. STRUCTURE analysis and principal coordinate analysis (PCoA) were used to investigate the patterns of population structure. The STRUCTURE analysis was performed using a Bayesian model-based clustering algorithm implemented in STRUCTURE version 2.3.4 [31], in which the admixture model with the correlated allele frequencies was used. A total of three independent replicates were run for each genetic cluster (K) value (K = 1–12), using a burn-in period of 100,000 iterations and 100,000 Markov Chain Monte Carlo (MCMC) repetitions. LnP(D) values were derived for each K and plotted to find the plateau of the  $\Delta K$  [32]. The probability of membership (Q value) equal to or greater than 0.70 was taken as a threshold to assign genotypes to a particular subpopulation, and those accessions with Q < 0.70 were considered genetically admixed. PCoA analysis was performed using DARwin 6.0.21 [33]. The final result was graphed using the ggplot2 R package [34].

### 2.5. Analysis of Population Differentiation and Analysis of Molecular Variance (AMOVA)

The number of subpopulations determined on the basis of the STRUCTURE analysis was used for AMOVA and pairwise  $F_{ST}$  analyses. The analysis of molecular variance (AMOVA) was performed using GenAlEx 6.51b2 [35] to calculate the sum of squares and variance components within and between populations.

## 3. Results

### 3.1. Sequencing and SNP Variant Detection in 50 Longan Accessions

In this study, we used the genotyping-by-sequencing (GBS) approach to identify genome-wide SNPs for longan. The ApeKI GBS libraries of 50 longan accessions generated the raw GBS data in each longan accession, ranging from 1.36 million to 6.76 million reads, with an average of 4.40 million reads (Table 1). Total nucleotides ranged from 0.20 to 1.00 Gb, with an average of 0.65 Gb. After filtering out low-quality reads, the total number of clean reads contained in the 50 libraries ranged from 1.35 million to 6.73 million, with an average of 4.38 million. The total number of nucleotides ranged from 0.20 to 0.99 Gb, with an average of 0.64 Gb, corresponding to 1.36-fold coverage of the longan genome (470 MB) (Table 1). The total number of filtered reads in all 50 samples was ~219 million.

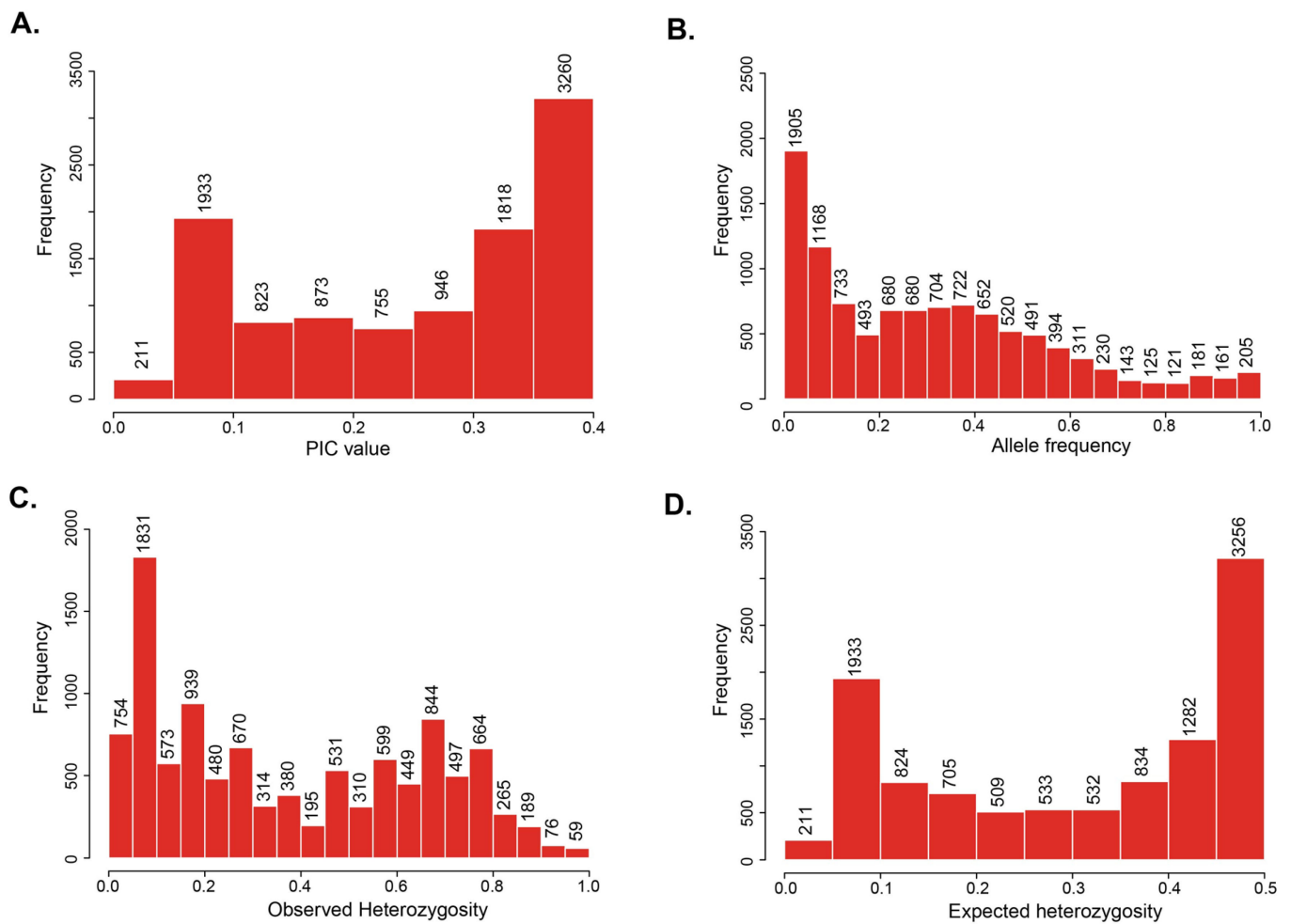
At the time of our experiments (2020–2021), there was no complete reference genome for longan. In this study, we instead aligned GBS reads to a draft genome of longan [26]. Using the GATK (Genome Analysis Toolkit) pipeline, we identified a total of 210,882 SNPs from the 50 longan accessions (Table S2). However, the miss rates of these SNPs in each longan accession were high, ranging from 45.65–71.94% (Table S2). We then further filtered these SNPs to obtain the non-missing SNPs and those with MAF > 2%. As a result, a total of 10,619 SNPs were obtained and used in further analyses (Table S3).

**Table 1.** Summary of sequencing reads generated for each of 50 longan accessions.

Accession	Name	Raw Reads (Million)	Raw Base (Gb)	Clean Reads (Million)	Clean Base (Gb)	Mapped Read (Million)	Average Read Depth
ACC01	Daw Kathi	3.73	0.55	3.71	0.55	3.16	23.86
ACC02	E-daw #1	6.05	0.89	6.03	0.89	2.92	23.92
ACC03	E-daw #2	4.81	0.71	4.79	0.71	2.35	21.59
ACC04	Biew Khiew #1	2.87	0.42	2.86	0.42	2.47	22.33
ACC05	Puang Thong #1	4.80	0.71	4.78	0.70	3.55	24.81
ACC06	E-daw #3	3.00	0.44	2.99	0.44	2.12	19.33
ACC07	E-daw #4	2.78	0.41	2.76	0.41	2.21	21.51
ACC08	E-daw #5	5.31	0.78	5.29	0.78	4.51	30.96
ACC09	Biew Khiew #2	4.88	0.72	4.85	0.72	4.23	32.87
ACC10	Kaew Yee	6.76	1.00	6.73	0.99	5.73	36.61
ACC11	E-daw #6	5.74	0.85	5.72	0.84	3.57	24.96
ACC12	E-daw #7	4.00	0.59	3.98	0.59	2.99	23.90
ACC13	E-daw #8	3.95	0.58	3.93	0.58	3.32	25.29
ACC14	Kohala	4.54	0.67	4.52	0.66	3.52	24.13
ACC15	E-daw #9	3.85	0.57	3.84	0.56	3.24	25.09
ACC16	E-daw #10	4.63	0.68	4.61	0.68	3.04	24.67
ACC17	E-daw #11	4.98	0.73	4.96	0.73	4.03	28.30
ACC18	E-daw #12	5.49	0.81	5.46	0.80	4.36	31.90
ACC19	E-daw #13	4.85	0.71	4.83	0.71	3.80	27.24
ACC20	E-daw #14	3.50	0.51	3.49	0.51	3.03	23.44
ACC21	Biew Khiew #3	4.86	0.71	4.84	0.71	4.04	30.51
ACC22	E-daw #15	3.61	0.53	3.60	0.53	3.08	23.59
ACC23	E-daw #16	4.04	0.59	4.03	0.59	3.31	24.92
ACC24	Chompoo #1	5.56	0.82	5.53	0.81	4.37	29.65
ACC25	Chuliang	3.73	0.55	3.71	0.55	2.61	24.68
ACC26	E-daw #17	5.78	0.85	5.75	0.85	4.80	32.14
ACC27	Krob Kathi #1	4.44	0.65	4.42	0.65	3.71	27.91
ACC28	E-daw #18	4.03	0.59	4.01	0.59	2.91	21.74
ACC29	Biew Khiew #4	6.11	0.90	6.08	0.89	5.25	32.76
ACC30	E-daw #19	4.40	0.64	4.38	0.64	3.55	25.99
ACC31	Puang Thong #2	4.00	0.59	3.98	0.58	3.04	23.28
ACC32	Krob Kathi #2	1.36	0.20	1.35	0.20	1.05	11.74
ACC33	Chompoo #2	3.69	0.54	3.68	0.54	2.96	23.17
ACC34	Biew Khiew #5	4.58	0.67	4.55	0.67	2.30	21.46
ACC35	Haew Yod Daeng	3.80	0.56	3.78	0.55	3.03	23.01
ACC36	E-daw #20	3.12	0.46	3.11	0.45	2.65	21.84
ACC37	Biew Khiew #6	5.73	0.84	5.70	0.83	3.56	26.24
ACC38	Puang Thong #3	4.83	0.71	4.79	0.70	3.86	25.26
ACC39	Puang Thong #4	3.55	0.52	3.53	0.52	2.81	22.25
ACC40	Chompoo #3	2.97	0.44	2.96	0.43	2.40	19.92
ACC41	E-daw #21	4.00	0.59	3.98	0.58	3.47	26.19
ACC42	Krob Kathi #3	2.47	0.36	2.46	0.36	1.99	18.17
ACC43	Chompoo #4	4.61	0.68	4.59	0.67	3.82	26.45
ACC44	E-daw #22	4.67	0.68	4.65	0.68	3.89	27.07
ACC45	E-daw #23	4.21	0.61	4.19	0.61	1.56	13.09
ACC46	E-daw #24	5.65	0.83	5.63	0.82	1.80	14.52
ACC47	Chompoo #5	4.90	0.72	4.88	0.71	1.94	14.13
ACC48	E-daw #25	5.12	0.75	5.10	0.74	2.23	15.26
ACC49	Biew Khiew #7	6.30	0.92	6.27	0.91	2.55	18.53
ACC50	E-daw #26	3.49	0.51	3.47	0.51	1.57	11.11
Avg		4.40	0.65	4.38	0.64	3.17	23.87
Min		1.36	0.20	1.35	0.20	1.05	11.11
Max		6.76	1.00	6.73	0.99	5.73	36.61
Total		220.12	32.33	219.09	32.17	158.29	1193.27

### 3.2. SNP Marker Properties

To explain the overall variability of each SNP marker in the 50 longan accessions, polymorphic information content (PIC) and allele frequency were estimated. As a result, the PIC values of the 10,619 SNPs were found to range from 0.04 to 0.40, with an average of 0.25 (Figure 1A). Allele frequency ranged from 0.02 to 0.98, with an average of 0.31 (Figure 1B). The level of observed heterozygosity ( $H_o$ ) ranged from 0.00 to 1.00, with an average of 0.38 (Figure 1C), and expected heterozygosity ( $H_e$ ) ranged from 0.04 to 0.50, with an average of 0.30 (Figure 1D).



**Figure 1.** Distribution of genetic diversity of 10,619 SNPs in the 50 longan accessions. (A) Polymorphic information content (PIC). (B) Allele frequency. (C) Observed heterozygosity (Ho). (D) Expected heterozygosity (He).

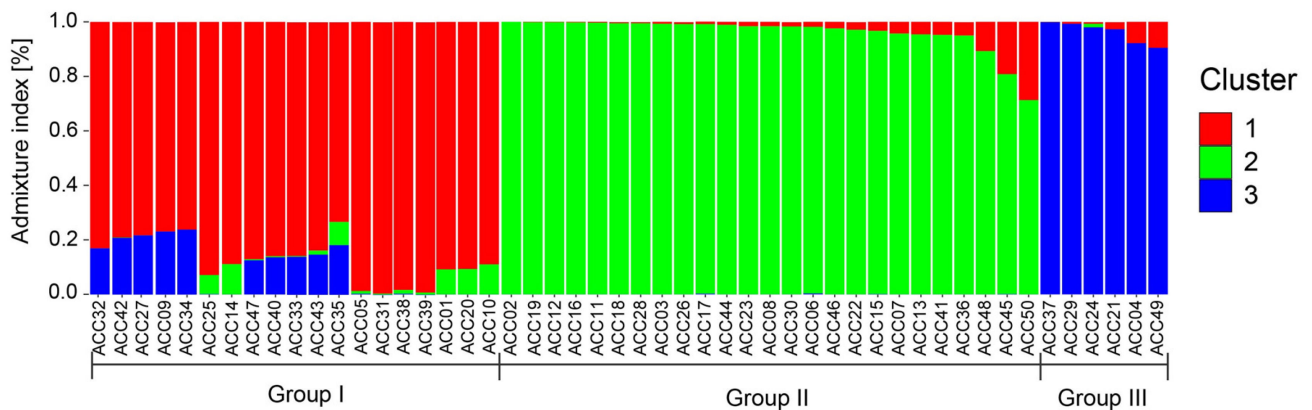
### 3.3. Population Structure and Genetic Relationships of Thai Longan Cultivars

A Bayesian clustering analysis in STRUCTURE (v.2.3.4) was performed to determine the population structure in the panel of 50 longan genotypes. The number of clusters (K) was plotted against the delta K to determine the appropriate K value. The assessed log likelihood [LnP(D)] with the delta K suggested the best clustering at K = 4 (Figure S1). Based on a probability threshold (Q) of 0.70, the 50 longan accessions were divided into four subpopulations (groups G1–4) and one admixture group (Figure 2B). Of these, group G1 consisted of 25 accessions, all of which were ‘E-daw’ cultivars (Table S4). The proportion of membership in most accessions in this group was greater than 90%, except for ACC45 (E-daw #23) and ACC50 (E-daw #26), which had Q values of 83.9% and 75.1%, respectively. Group G2 consisted of seven accessions, of which four were ‘Puang Thong’ cultivars (ACC05, ACC31, ACC38, and ACC39), and the other three were ‘Daw Kathi’ (ACC01), ‘E-daw’ (ACC20), and ‘Kaew Yee’ (ACC10). The Q values in these groups ranged from 81.6% to 99.9%. Group G3 consisted of six accessions, five of which were ‘Biew Khiew’ cultivars (ACC37, ACC29, ACC21, ACC04, and ACC49), and one of which was a ‘Chompoo’ cultivar (ACC24). The Q values in this group ranged from 91.0% to 100%. Group G4 contained five accessions, of which three were ‘Krob Kathi’ cultivars (ACC27, ACC32, and ACC42) with Q values above 98%, and the other two accessions were ‘Biew Khiew’ cultivars (ACC09 and ACC34) with Q values of 83.9% and 83.3%, respectively. There were also seven accessions with admixed genotypes. Among them, there were

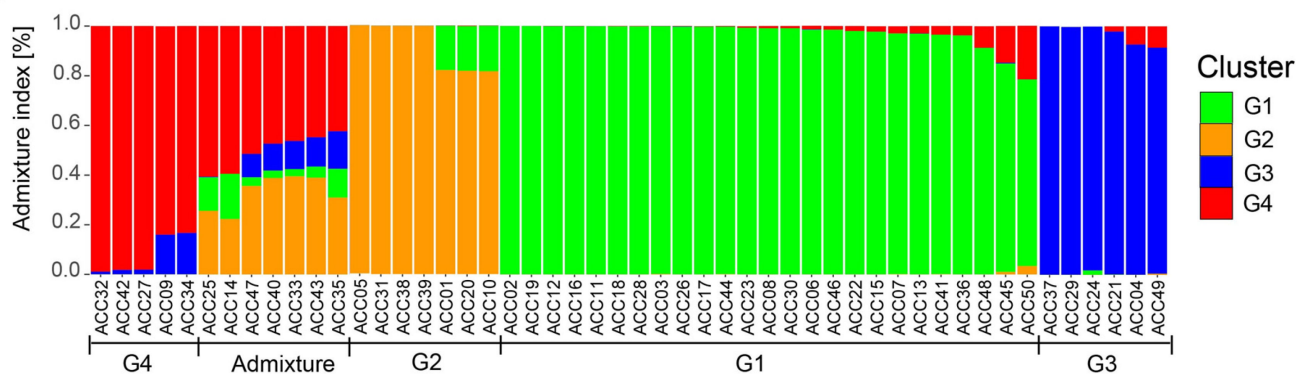


five accessions of Thai longan, four of which were ‘Chompoo’ cultivars (ACC33, ACC43, ACC40, and ACC47), and another was the ‘Haew Yod Daeng’ cultivar (ACC35). All Thai longan accessions in this admixture group contained mixed genotypes from all four groups in different proportions. The other two accessions within the admixture group were the two foreign cultivars, Chuliang (ACC25) and Kohala (ACC14), which contained admixed genotypes of groups G1, G2, and G4 (Figure 2).

A.

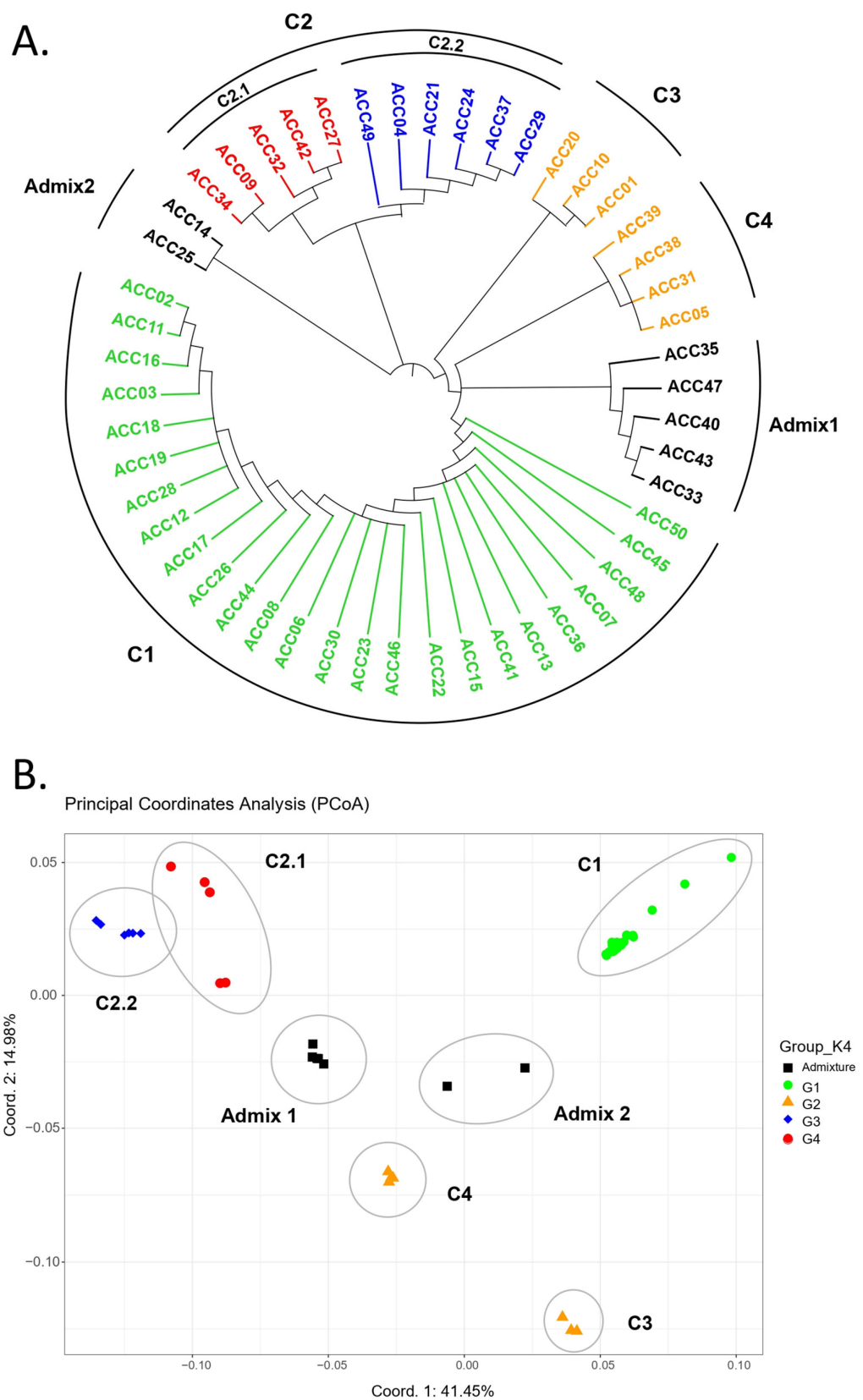


B.



**Figure 2.** Population structure analysis of the panel of longan genotypes inferred using STRUCTURE software for (A) delta K = 3 and (B) delta K = 4. Stack bars represent samples. Each color represents the highest ancestry of the cluster. Individuals with multiple colors have admixed genotypes from multiple clusters. Color blocks represent each subpopulation.

Furthermore, we analyzed the genetic relationships among the 50 longan accessions using UPGMA-based clustering analysis and principal coordinate analysis (PCoA) based on the 10,619 SNPs. The UPGMA tree revealed six main clusters (C1–C4, Admix1, and Admix 2) of the 50 longan accessions (Figure 3A; Table S4). Group C1 (25 accessions) was similar to group G1, as identified by STRUCTURE. Group C2 (11 accessions) contained two subgroups, C2.1 and C2.2, similar to groups G2 and G3, respectively, as determined by STRUCTURE. Groups C3 (4 accessions) and C4 (3 accessions) together resembled group G4 from STRUCTURE. The groups Admix1 and Admix2 belonged to the admixture group of STRUCTURE. The admixture groups of Thai longan accessions (ACC35, ACC47, ACC40, ACC43, and ACC33) and those of foreign longan varieties (ACC14 and ACC25) were clearly separated (Figure 3A). PCoA analysis revealed longan clusters consistent with STRUCTURE and the UPGMA tree (Figure 3B; Table S4). The first two PCs explained 41.45% and 14.98% of the variation, respectively.



**Figure 3.** Phylogenetic tree and PCoA for 50 longan accessions. **(A)** UPGMA phylogenetic tree. Different colors in the plots indicate different STRUCTURE groups: green indicates G1; orange indicates G2; blue indicates G3; and red indicates G4. **(B)** Two-dimensional plots of PCoA (coordinate 1 vs. coordinate 2). Each circle represents a group corresponding to the UPGMA clusters.

### 3.4. Genetic Differentiation of Populations

We defined clusters of the 50 longan accessions based on the four groups of STRUC-TURE (groups G1–4) and analyzed the genetic differentiation of the populations using an analysis of molecular variance (AMOVA) and Wright's  $F$  statistics. Seven accessions with admixed genotypes were excluded from the analysis. The AMOVA for the four subpopulations (43 accessions) indicated that 74% of the variation was caused by differences among groups, while the remaining 26% was caused by differences within groups (Table 2).  $H_o$  and  $H_e$  in the four groups ranged from 0.34 to 0.44 and from 0.22 to 0.32, respectively. The average genetic differentiation coefficient ( $F_{ST}$ ) among the four populations was 0.25, indicating a considerable degree of differentiation among the four subpopulations (Table 3). According to the results of the pairwise  $F_{ST}$  analysis, populations G2 and G4 were the most closely related ( $F_{ST} = 0.25$ ), whereas populations G1 and G3 showed a substantial degree of differentiation ( $F_{ST} = 0.40$ ) (Table 4).  $F_{IS}$  values were negative in each subpopulation, indicating the absence of inbreeding.

**Table 2.** Summary of AMOVA for the 43 longan accessions.

Source of Variation	df	SS	MS	Variance Components	Percentage of Total Variance	Probability ( $p$ )
Among Pops	3	48,126.59	16,042.19	1784.43	74%	<0.001
Within Pops	39	24,667.50	632.50	632.50	26%	
Total	42	72,794.09		2416.93	100%	

**Table 3.** Statistics of genetic variation for the 43 longan accessions.

Population/Cluster	No. of Accessions	PIC	$H_o$	$H_e$	$F_{IS}$	$F_{ST}$
G1	25	0.36	0.44	0.24	−0.79	0.25
G2	7	0.32	0.42	0.32	−0.32	
G3	6	0.36	0.4	0.22	−0.81	
G4	5	0.32	0.34	0.23	−0.42	

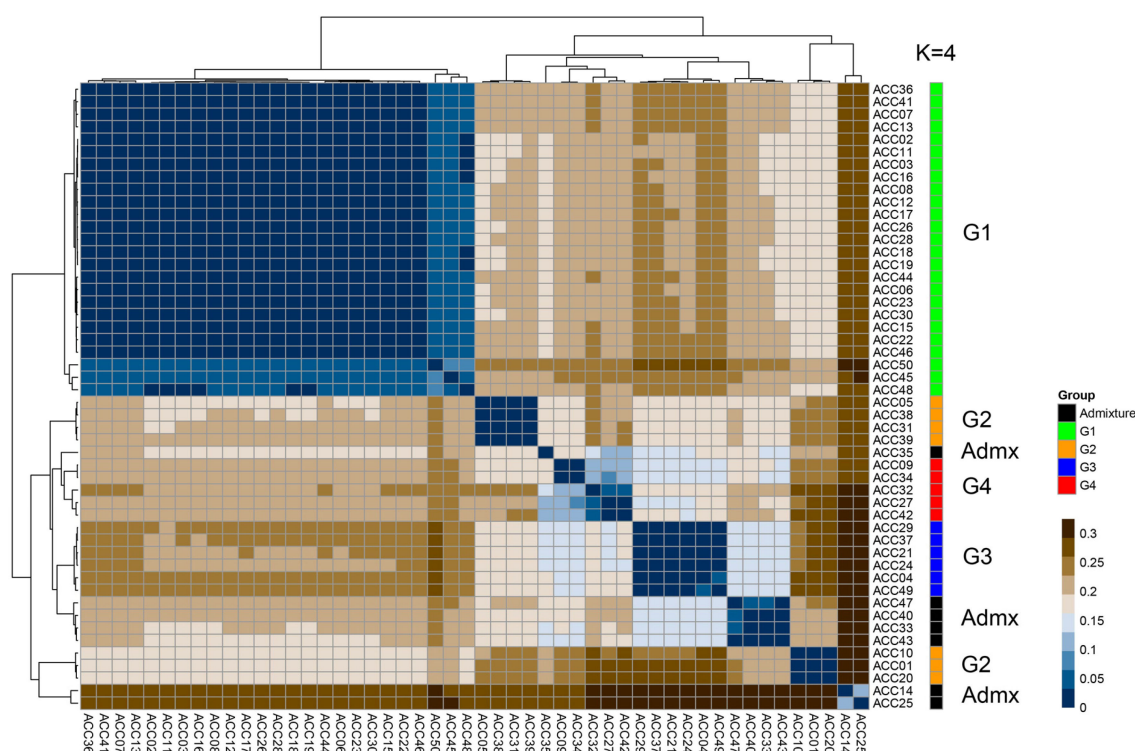
**Table 4.** Pairwise genetic differentiation values ( $F_{ST}$ ) of 43 longan accessions.

Population/Cluster	G1	G2	G3	G4
G1	0			
G2	0.258	0		
G3	0.403	0.267	0	
G4	0.359	0.246	0.291	0

### 3.5. Genetic Distance of the 50 Longan Accessions

The observed heterozygosity ( $H_o$ ) in the 50 accessions based on the 10,619 SNPs ranged from 19.27% (ACC14: Kohala) to 44% (ACC02: E-daw #1) (Table S3). Most of the longan accessions in this study had a  $H_o$  value greater than 25%. We determined the relatedness among the 50 longan accessions, which were classified into four groups (G1–4 and admixture) based on the STRUC-TURE analysis results. The similarity among accessions within each subpopulation was relatively high (Figure 4). The accessions in the G2 group of STRUC-TURE were divided into two subgroups in accordance with the UPGMA tree analysis results. It was clearly shown that the two foreign longan accessions (ACC14 and ACC25) were distantly related to all Thai longan accessions.





**Figure 4.** Hierarchical cluster dendrogram and heatmap based on the dissimilarity matrix from 10,619 SNPs representing the genetic relationships among 50 longan accessions. The four STRUC-TURE subpopulations and admixtures are shown with G1-4 and Admx and color bars corresponding to each STRUC-TURE group. Degrees of relatedness are indicated by colors ranging from dark blue (strong relatedness) to brown (no relatedness).

#### 4. Discussion

Thailand is one of the major longan producers and the largest exporter. Longan production in Thailand is centered in the northern areas, i.e., Chiang Mai, Lamphun, and Chiang Rai provinces (Office of Agricultural Economics, 2021: <https://www.oae.go.th>, accessed date: 1 May 2023), where elevations range from 300 to 600 m above sea level and winters are cool. Production is also found in eastern areas, i.e., in Chanthaburi province (Office of Agricultural Economics, 2021: <https://www.oae.go.th>, accessed date: 1 May 2023). More than 20 longan cultivars are grown in Thailand. The most common cultivars grown commercially in the north include E-daw, Biew Khiew, Chompoo (or Si Chompoo), and Heaw. E-daw is the most popular longan cultivar in Thailand because it flowers more readily and fruits more regularly than other cultivars, accounting for about 73% of plantings [6]. It flowers in mid-January, and its fruits are harvested from July to August. Biew Khiew is a late bloomer and blooms in alternate years. In the bearing year, it flowers in late January, and its fruits are harvested in late August. Chompoo is a medium bearer and medium flowerer, but with an irregular fruit set. It flowers in late January, and its fruit is harvested from late July to August. Haew is a medium fruit bearer and tends to have a variable fruit set. It flowers in late January, and its fruits are harvested from July to August [36]. Commercial longan cultivars are usually propagated by vegetative propagation.

In plant breeding, the availability of genetic diversity is a key factor in the improvement of a crop. Moreover, accurate variety identification is essential for germplasm management and breeding. Recently, a phylogenetic study of the genus *Dimocarpus*, which includes 26 accessions of Thai longan (*D. longan* var. *longan*), revealed that longan cultivars with a similar name may have different genetic backgrounds [5]. The genetic diversity and population structure of longan have been studied in detail using Chinese longan germplasm,

but only a few longan cultivars from Thailand have been included [2,16]. Therefore, the genetic diversity and population structure of Thai longans still need to be explored. In this study, we used GBS technology to identify 10,619 high-quality SNPs in a collection of 50 longan accessions obtained from a community biobank. GBS can be carried out in species with or without reference genome sequences [37]. However, a high-quality reference genome sequence would greatly facilitate the discovery of high-density SNP markers. At the time of our experiments (2020–2021), only the draft genome of longan was available [16]. Therefore, we based our analysis on this draft genome. In our analysis based on the 10,916 SNPs, the mean PIC value of 0.25 indicated that the 50 longan accessions exhibited moderate polymorphism [38]. The low PIC value of the SNP markers could be due to their biallelic nature [39]. A model-based population structure analysis divided the population into five subgroups, most of which overlapped with cultivar names. However, there were also several accessions with the same cultivar names distributed among different groups. We suspect that this may be due to a mislabeling of the accessions in the biobank rather than divergence of accessions with the same cultivar name. The AMOVA results showed that a high level of genetic diversity was observed among subpopulations. This result reflected the asexual reproduction of most longan cultivars in Thailand. Accessions with high similarity, i.e., E-daw longan cultivars in STRUCTURE group G1, may reflect the clonality of these cultivars because commercial longan cultivars are usually propagated by asexual reproduction.

Based on 60 SNP markers developed from transcriptome sequences, Wang et al. identified two subclusters for 25 cultivated longan accessions [16]. Thai cultivars and other cultivars from Hawaii used in the study were clearly separated from Chinese cultivars. In addition, based on the results of phylogenetic and PCA analyses using 1,421,213 high-quality SNP loci, 87 Chinese longan accessions were classified into Guangdong and Fujian groups [2]. Based on these studies, Thai longan cultivars have been found to be more closely related to western Guangdong cultivars but not to Fujian cultivars [2,16]. In this study, we included a Chinese longan cultivar, Chuliang, which is a Guangdong cultivar, and found that it has admixed genotypes shared with three subpopulations of Thai longan accessions. The same pattern of admixed genotypes was also found in Kohala longan from the USA. The Thai longan cultivars of STRUCTURE group G3 do not share genotypes with the Chinese longan, Chuliang, or the US longan, Kohala. However, more longan accessions from China and other countries are needed to evaluate the relationship between foreign longans and the Thai longan. The results of this study on longan diversity and population structure pave the way for breeding programs to select suitable parents to produce new longan varieties with better yield and quality.

## 5. Conclusions

Using GBS technology, we identified 210,882 SNPs from 50 longan accessions. A total of 10,916 high-quality SNPs were used to estimate genetic diversity and population structure among the 50 longan accessions. Four major subgroups were identified by STRUCTURE, UPGMA tree, and PCoA. Accessions within the same group were closely related, while accessions in different groups differed significantly. These results will be useful in the verification of longan cultivars as well as in managing longan germplasm and may facilitate the genetic improvement of longan.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/horticulturae9060726/s1>. Figure S1: the average log-likelihood of the K-value against the number of K and delta K for different numbers of subpopulations (K); Table S1: list of longan accessions used for GWAS analysis; Table S2: summary of 210,882 total SNPs identified based on GBS data in each longan accession; Table S3: summary of 10,619 filtered SNPs in each longan accession; Table S4: clusters of 50 longan accessions identified by STRUCTURE and UPGMA tree.

**Author Contributions:** S.A., S.W., V.R. and T.T. conceived and designed the experiments. C.S. performed the bioinformatics analysis. K.R. and R.D. performed lab experiments. S.W. and S.A. provided critical discussion. K.R., S.W. and M.K.P. wrote the manuscript. S.W. and S.A. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Biodiversity-based Economy Development Office (Public Organization)—BEDO (grant number BEDO-PS60194).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding authors.

**Acknowledgments:** We thank to the Nong Chang Khuen Longan Community BioBank, Lamphun, Thailand for providing longan plant materials.

**Conflicts of Interest:** The Authors declare no competing interest.

## References

1. Tindall, H.D. Sapindaceous fruits: Botany and horticulture. In *Horticultural Reviews*; Janick, J., Ed.; John Wiley & Sons, Inc.: Oxford, UK, 1994; pp. 143–196. ISBN 9780470650561.
2. Wang, J.; Li, J.; Li, Z.; Liu, B.; Zhang, L.; Guo, D.; Huang, S.; Qian, W.; Guo, L. Genomic insights into longan evolution from a chromosome-level genome assembly and population genomics of longan accessions. *Hortic. Res.* **2022**, *9*, uhac021. [[CrossRef](#)] [[PubMed](#)]
3. Li, J.; Miao, S.; Jiang, Y. Changes in quality attributes of longan juice during storage in relation to effects of thermal processing. *J. Food Qual.* **2009**, *32*, 48–57. [[CrossRef](#)]
4. Liu, G.; Sun, J.; He, X.; Tang, Y.; Li, J.; Ling, D.; Li, C.; Li, L.; Zheng, F.; Sheng, J.; et al. Fermentation process optimization and chemical constituent analysis on longan (*Dimocarpus longan* Lour.) wine. *Food Chem.* **2018**, *256*, 268–279. [[CrossRef](#)] [[PubMed](#)]
5. Lithanatudom, S.K.; Chaowasku, T.; Nantararat, N.; Jaroenkit, T.; Smith, D.R.; Lithanatudom, P. A first phylogeny of the genus *dimocarpus* and suggestions for revision of some taxa based on molecular and morphological evidence. *Sci. Rep.* **2017**, *7*, 6716. [[CrossRef](#)]
6. Menzel, C.; Waite, G.K. *Litchi and Longan: Botany, Production and Uses*, 1st ed.; CABI: Wallingford, Oxfordshire, UK, 2005; p. 336, ISBN 0851996965.
7. Sakata, S. *New Trends and Challenges for Agriculture in the Mekong Region: From Food Security to Development of Agri-Businesses*; JETRO Bangkok/IDE-JETRO; Bangkok Research Center: Bangkok, Thailand, 2019.
8. Subhadrabandhu, S.; Yapwattanaphun, C. Lychee and longan production in thailand. *Acta Hortic.* **2001**, *558*, 49–57. [[CrossRef](#)]
9. Said, S.; Bayu Putra, W.P.; Anwar, S.; Agung, P.P.; Yuhani, H. Phenotypic, morphometric characterization and population structure of Pasundan cattle at West Java, Indonesia. *Biodiversitas* **2017**, *18*, 1638–1645. [[CrossRef](#)]
10. Nadeem, M.A.; Nawaz, M.A.; Shahid, M.Q.; Doğan, Y.; Comertpay, G.; Yıldız, M.; Hatipoğlu, R.; Ahmad, F.; Alsaleh, A.; Labhane, N.; et al. DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* **2017**, *32*, 261–285. [[CrossRef](#)]
11. Lin, T.; Lin, Y.; Ishiki, K. Genetic diversity of *Dimocarpus longan* in China revealed by AFLP markers and partial *rbcl* gene sequences. *Sci. Hortic.* **2005**, *103*, 489–498. [[CrossRef](#)]
12. Mei, Z.Q.; Fu, S.Y.; Yu, H.Q.; Yang, L.Q.; Duan, C.G.; Liu, X.Y.; Gong, S.; Fu, J.J. Genetic characterization and authentication of *Dimocarpus longan* Lour. using an improved RAPD technique. *Genet. Mol. Res.* **2014**, *13*, 1447–1455. [[CrossRef](#)]
13. Ho, V.T.; Ngo, Q.N. Development of SCAR makers for longan (*Dimocarpus longan* L.) authentication in Vietnam. *BioTechnologia* **2018**, *99*, 401–407. [[CrossRef](#)]
14. Zhou, J.; Fu, J.-x.; Wu, Z.-x.; Huang, S.-s.; Zhang, Y.-f.; Wang, Y.; Hu, Y.-l.; Hu, G.-b.; Liu, C.-m. Genetic diversity in litchi and longan germplasm as determined by srp markers. *Acta Hortic.* **2011**, *918*, 799–805. [[CrossRef](#)]
15. Fu, J.-x.; Wang, Y.; Zhou, J.; Zhao, H.-y.; Huang, S.-s.; Hu, Y.-l.; Hu, G.-b.; Liu, C.-m. Genetic diversity of germplasm resources of litchi and longan using ssr analysis. *Acta Hortic.* **2011**, *918*, 363–370. [[CrossRef](#)]
16. Wang, B.; Tan, H.-W.; Fang, W.; Meinhardt, L.W.; Mischke, S.; Matsumoto, T.; Zhang, D. Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences for identification of longan (*Dimocarpus longan*) germplasm. *Hortic. Res.* **2015**, *2*, 14065. [[CrossRef](#)] [[PubMed](#)]
17. Kumar, S.; Banks, T.W.; Cloutier, S. SNP Discovery through Next-Generation Sequencing and Its Applications. *Int. J. Plant Genom.* **2012**, *2012*, 831460. [[CrossRef](#)]
18. Davey, J.W.; Hohenlohe, P.A.; Etter, P.D.; Boone, J.Q.; Catchen, J.M.; Blaxter, M.L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **2011**, *12*, 499–510. [[CrossRef](#)]
19. Vasumathy, S.K.; Peringottillam, M.; Sundaram, K.T.; Kumar, S.H.K.; Alagu, M. Genome- wide structural and functional variant discovery of rice landraces using genotyping by sequencing. *Mol. Biol. Rep.* **2020**, *47*, 7391–7402. [[CrossRef](#)]
20. Gouesnard, B.; Negro, S.; Laffray, A.; Glaubitz, J.; Melchinger, A.; Revilla, P.; Moreno-Gonzalez, J.; Madur, D.; Combes, V.; Tollon-Cordet, C.; et al. Genotyping-by-sequencing highlights original diversity patterns within a European collection of 1191 maize flint lines, as compared to the maize USDA genebank. *Theor. Appl. Genet.* **2017**, *130*, 2165–2189. [[CrossRef](#)]

21. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [\[CrossRef\]](#)
22. Alipour, H.; Bihamta, M.R.; Mohammadi, V.; Peyghambari, S.A.; Bai, G.; Zhang, G. Genotyping-by-Sequencing (GBS) Revealed Molecular Genetic Diversity of Iranian Wheat Landraces and Cultivars. *Front. Plant Sci.* **2017**, *8*, 1293. [\[CrossRef\]](#)
23. Bastien, M.; Boudhrioua, C.; Fortin, G.; Belzile, F. Exploring the potential and limitations of genotyping-by-sequencing for SNP discovery and genotyping in tetraploid potato. *Genome* **2018**, *61*, 449–456. [\[CrossRef\]](#)
24. Riengwong, K.; Wanchana, S.; Aesomnuk, W.; Saensuk, C.; Nubankoh, P.; Ruanjaichon, V.; Kraithong, T.; Toojinda, T.; Vanavichit, A.; Arikrit, S. Mining and validation of novel genotyping-by-sequencing (GBS)-based simple sequence repeats (SSRs) and their application for the estimation of the genetic diversity and population structure of coconuts (*Cocos nucifera* L.) in Thailand. *Hortic. Res.* **2020**, *7*, 156. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Lin, Y.; Min, J.; Lai, R.; Wu, Z.; Chen, Y.; Yu, L.; Cheng, C.; Jin, Y.; Tian, Q.; Liu, Q.; et al. Genome-wide sequencing of longan (*Dimocarpus longan* Lour.) provides insights into molecular basis of its polyphenol-rich characteristics. *Gigascience* **2017**, *6*, 1–14. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [\[CrossRef\]](#)
28. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *11*, 11.10.1–11.10.33. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Jombart, T. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403–1405. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Kumar, S.; Stecher, G.; Li, M.; Nnyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [\[CrossRef\]](#) [\[PubMed\]](#)
31. 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. [\[CrossRef\]](#)
32. Earl, D.A.; vonHoldt, B.M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [\[CrossRef\]](#)
33. Perrier, X. DARwin Software. 2006. Available online: <http://darwin.cirad.fr/darwin> (accessed on 1 May 2023).
34. Wickham, H. *ggplot2—Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-0-387-98140-6.
35. Peakall, R.O.D.; Smouse, P.E. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [\[CrossRef\]](#)
36. Ghosh, S.N. *Breeding of Underutilized Fruit Crops*; JAYA Publishing House: Delhi, India, 2015; ISBN 9384337404.
37. Kagale, S.; Koh, C.; Clarke, W.E.; Bollina, V.; Parkin, I.A.P.; Sharpe, A.G. Analysis of Genotyping-by-Sequencing (GBS) Data. *Methods Mol. Biol.* **2016**, *1374*, 269–284. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331. [\[PubMed\]](#)
39. Mammadov, J.; Aggarwal, R.; Buyyarapu, R.; Kumpatla, S. SNP markers and their impact on plant breeding. *Int. J. Plant Genom.* **2012**, *2012*, 728398. [\[CrossRef\]](#) [\[PubMed\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.