

Article

Genetic Mapping by Sequencing More Precisely Detects Loci Responsible for Anaerobic Germination Tolerance in Rice

John Carlos I. Ignacio^{1,2,3,4}, Maricris Zaidem^{3,5}, Carlos Casal, Jr.¹, Shalabh Dixit¹, Tobias Kretzschmar^{1,6}, Jaime M. Samaniego², Merlyn S. Mendioro², Detlef Weigel³ and Endang M. Septiningsih^{1,7,*}

¹ International Rice Research Institute, DAPO Box 7777, Metro Manila 1301, Philippines; ignacio.8@osu.edu (J.C.I.I.); c.casal@irri.org, (C.C.J.); s.dixit@irri.org (S.D.); Tobias.Kretzschmar@scu.edu.au (T.K.)

² University of the Philippines, Los Baños, Laguna 4031, Philippines; jmsamaniego2@up.edu.ph (J.M.S.); msmendioro@up.edu.ph (M.S.M.)

³ Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany; mzaidem@gmail.com (M.Z.); detlef.weigel@tuebingen.mpg.de (D.W.)

⁴ Department of Horticulture and Crop Science, Ohio Agriculture Research and Development Center, The Ohio State University, 1680 Madison Ave, Wooster, OH 44691, USA

⁵ Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK

⁶ Southern Cross Plant Sciences, Southern Cross University, 1 Military Road, Lismore, NSW 2480, Australia

⁷ Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843, USA

* Correspondence: esepiningsih@tamu.edu



Citation: Ignacio, J.C.I.; Zaidem, M.; Casal, C.J.; Dixit, S.; Kretzschmar, T.; Samaniego, J.M.; Mendioro, M.S.; Weigel, D.; Septiningsih, E.M. Genetic Mapping by Sequencing More Precisely Detects Loci Responsible for Anaerobic Germination Tolerance in Rice. *Plants* **2021**, *10*, 705. <https://doi.org/10.3390/plants10040705>

Academic Editor: Chih-Wei Tung

Received: 6 March 2021

Accepted: 4 April 2021

Published: 6 April 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: Direct seeded rice (DSR) is a mainstay for planting rice in the Americas, and it is rapidly becoming more popular in Asia. It is essential to develop rice varieties that are suitable for this type of production system. ASD1, a landrace from India, possesses several traits desirable for direct-seeded fields, including tolerance to anaerobic germination (AG). To map the genetic basis of its tolerance, we examined a population of 200 F_{2,3} families derived from a cross between IR64 and ASD1 using the restriction site-associated DNA sequencing (RAD-seq) technology. This genotyping platform enabled the identification of 1921 single nucleotide polymorphism (SNP) markers to construct a high-resolution genetic linkage map with an average interval of 0.9 cM. Two significant quantitative trait loci (QTLs) were detected on chromosomes 7 and 9, *qAG7* and *qAG9*, with LOD scores of 7.1 and 15.0 and R² values of 15.1 and 29.4, respectively. Here, we obtained more precise locations of the QTLs than traditional simple sequence repeat and low-density SNP genotyping methods and may help further dissect the genetic factors of these QTLs.

Keywords: quantitative trait locus (QTL); sequencing based genotyping; RAD-seq; anaerobic germination (AG); rice (*Oryza sativa*)

1. Introduction

Farmers have increasingly adopted direct seeded rice (DSR) in both rainfed and irrigated ecosystems since it requires less labor and cost of production compared to conventional transplanting [1]. However, unlevelled fields or heavy rainfall on direct-seeded fields can lead to poor crop establishment as most rice varieties are susceptible to flooding during germination or anaerobic germination (AG). Hence, breeding for DSR requires tolerance to AG stress and is essential to sustain rice production in these ecosystems. Screening of over 8000 gene bank accessions and breeding lines at the International Rice Research Institute (IRRI) identified only 19 tolerant lines with greater than 70% survival rate [2]. Additional efforts to screen more varieties had been conducted at IRRI (E. Septiningsih, unpublished data). Some of these accessions have been used for physiological and agronomical studies, and as donors in genetic studies and crop improvement efforts [2–17].

Several biparental quantitative trait loci (QTL) mapping studies for AG tolerance have been performed previously using various genotyping platforms. For example, a

study by Jiang et al. [18] reported five QTLs for AG using restriction fragment length polymorphism (RFLP) markers on a mapping population derived from Kinmaze/DV85. Using 135 simple-sequence repeat (SSR) markers, Angaji et al. [2] identified five QTLs, including *qAG-9-2* or *AG1*, on a BC₂F₂ population derived from IR64/Khao Hlan On. *AG1* has been used in breeding efforts (Toledo et al. 2015) and the gene underlying this QTL had been cloned as trehalose-6-phosphate phosphatase (*OsTPP7*) [14]. Six QTLs were identified in a study by Septiningsih et al. [11], including a major QTL on chromosome 7, derived from an F_{2:3} population of IR42/Ma-Zhan Red, using 118 SSR markers. Baltazar et al. [12] identified three QTLs, including the major one on chromosome 7, derived from an F_{2:3} population of IR64/Nanhi, using 384-plex single nucleotide polymorphism (SNP) Indica/Indica set on the Illumina BeadXpress Reader, resulting of 234 SNP markers used in the map. Additionally, using the same genotyping platform but with an F_{2:3} population derived from IR64/Kharsu 80A, resulting in 217 polymorphic markers, Baltazar et al. [13] identified four QTLs, of which three of them were detected on chromosome 7. More recently, Ghosal et al. [4] identified five QTLs on a BC₁F_{2:3} mapping population of NSIC Rc238/Kalarata using 185 kompetitive allele specific PCR (KASP) SNP markers. Using the same genotyping platform of 189 KASP SNP markers and the same donor, Kalarata, but a different elite cultivar, NSIC Rc222, the team identified four QTLs using a BC₁F_{2:3} mapping population. Three common QTLs on chromosomes 3, 6, and 7, were detected across the two populations. Additionally, Ghosal et al. [9] also reported four AG QTLs from a BC₁F_{2:3} mapping population derived from NSIC Rc222/BJ1 through bulk-segregant analysis strategy using 102 SSR markers. While these studies have pointed to a number of significant QTLs for AG tolerance in rice, their limited marker resolution has made it challenging to narrow down the QTL regions.

Exploration of simple, inexpensive high-density scalable single nucleotide polymorphism (SNP) genotyping led to the development of sequencing of restriction site-associated DNA (RAD) tags [19]. This RAD sequencing (RAD-seq) approach utilizes one or two restriction enzymes to digest genomic DNA samples at many short identical sites so that the resulting fragments can be uniquely barcoded. Barcoding allows samples to be multiplexed in a library and genotyped together in a single sequencing run. Elshire et al. [20] have also reported a similar approach called Genotyping-by-Sequencing (GBS). Sequencing-based genotyping is so robust that it can be used in a wide range of genetic studies, such as genome-wide association studies (GWAS), genomic selection (GS), diversity studies, and genome mapping, even without prior knowledge of the species [19,21]. QTL mapping using RAD-seq as a genotyping platform has been performed in various crops, including soybean [22–25], peanut [26–28], cowpea [29], bitter melon [30,31], Chinese cabbage [32], alfalfa [33], sesame [34], foxtail millet [33,35], sorghum [36], and rice [37,38].

ASD1, an *Indica* rice variety from India [39], has been characterized for several traits desired for DSR, including AG tolerance, and found to be a promising genetic donor for this trait [40]. In this study, we aim to map the genetic factors responsible for AG tolerance from ASD1 with the use of a biparental F₂ population and RAD-seq technology as the genotyping platform.

2. Results

2.1. Phenotypic Performance

From the 418 randomly selected families and parental lines tested for germination rate under normal conditions, 118 had germination rates of 60 to 97%, while 300 families had germination rates higher than 97%, which were then tested for survivability under AG. The parental controls also showed 97–100% germination rates under normal conditions. Among the F_{2:3} families, the survival rates under AG were ranging between 0% and 83%. The average survival rates of the parents, IR64 and ASD1, under AG were 11.67% and 46.55%, respectively (Figure 1; Figure S1). The mapping population demonstrated transgressive segregation in both directions, where 60 out of 300 families (20%) had lower survival rates than that of IR64, while 50 out of the 300 families (17%) had higher survival rates than

that of ASD1. IR64 is considered moderately sensitive to AG compared to other sensitive accessions [2], and, even in the case of mapping populations made with a highly sensitive accession, such as IR42, transgressive segregation could still be observed [11].

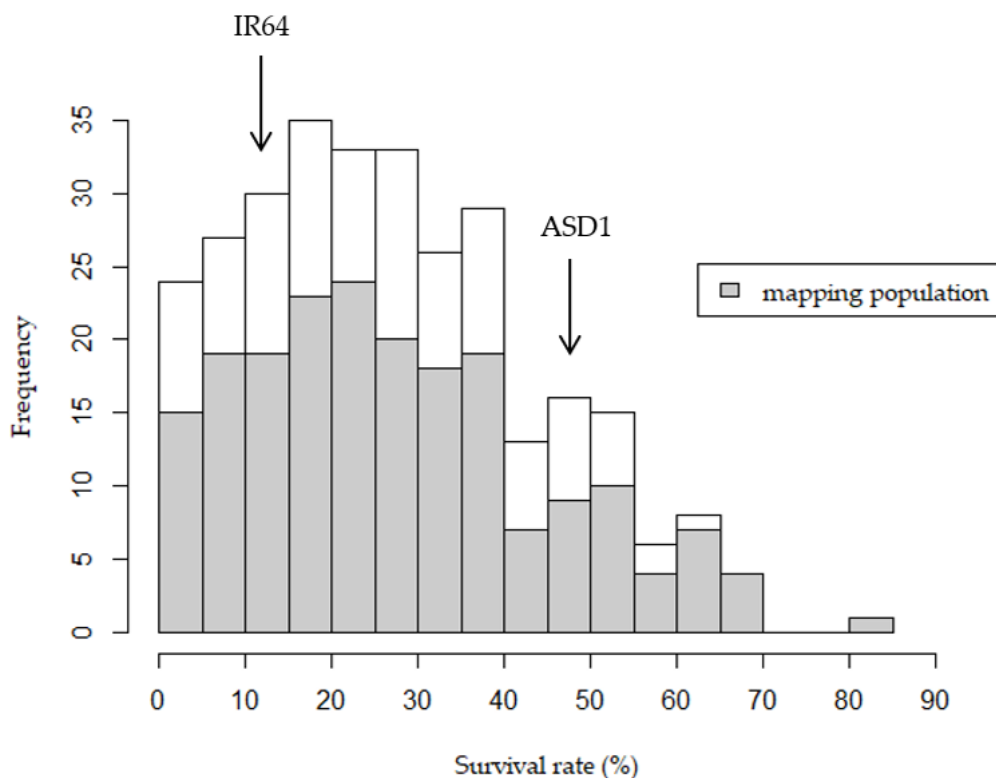


Figure 1. Survival rates of the 300 IR64 × ASD1 F_{2.3} families under anaerobic germination, along with the average survival of the parents. The 199 families used in quantitative trait loci (QTL) mapping are shown in grey.

2.2. Genotyping

2.2.1. Library Preparation and Sequencing

All extracted DNA samples had 260/230 absorbance ranging from 1.75 to 1.85 and had 260/230 absorbance of 1.8 to 2.2. Based on the Agilent 2100 Bioanalyzer DNA 1000 chip results, each of the libraries had concentrations of around 300 pg/mL and fragment size ranging 300–600 bp. Each sequencing lane (also each library) produced roughly 380 million reads with length of 101 bp and an average GC content of 45%.

2.2.2. Variant Detection

For each of the three raw sequencing outputs, about 200 million duplicated reads were removed. About 80% of the total reads were demultiplexed, and the average read count was 1.54 million per sample, with only 32 samples having less than 100,000 reads. The combination of the 3 replicated parents from all sequencing lanes resulted in 3.02 million reads assigned to IR64, while 4.19 million reads to ASD1. On average, about 97% of the deduplicated and demultiplexed reads from each sequencing lane aligned with the reference genome. A total of 156,636 variants were discovered in 270 F_{2.3} families and two parental lines. Among these, 137,819 were SNPs, and 18,817 were indels.

2.2.3. Variant and Sample Filtering

All 18,817 indels and 1659 multiallelic SNPs (allele count > 2) were filtered out of the dataset. Out of the 136,160 remaining SNPs, only 6068 were homozygous and identified to be polymorphic between the two parents. SNPs with less than 0.18 minimum allele frequencies and heterozygous proportions less than 0.05 and greater than 0.55 were excluded from the dataset, removing 321 sites. After removing SNPs from the same reads,

the number of SNPs was further reduced by 2007. By taking only samples and markers with call rates of at least 50% and 60%, respectively, the dataset had a remaining total of 200 samples and 3436 markers and was used for imputation and genotyping call correction.

2.2.4. Adapting Genotyping Data for F₂ Genetic Analysis

Missing data proportion was greatly reduced from 20.4% to 6.1% after imputation (Table S1). A closer to 1AA:2AB:1BB Mendelian genotype F₂ ratio was attained after correcting for under-called heterozygous genotypes. The remaining steps did not greatly affect the distributions of genotypes but influenced the estimation of the genetic map where initial estimated maps without correction reached more than 1000 cM per chromosome (data not shown). Furuta et al. [41] noted that, when GBS data on a rice F₂ population was left uncorrected, the genetic map of a chromosome can reach up to 3500 cM. In their case, correction reduced total map size of all chromosomes to 1536 cM, which was similar to our observation.

2.3. QTL Analysis

2.3.1. Genetic Map Estimation

Before map estimation, 1515 SNP markers contained duplicate information and were removed from the dataset. The map was estimated using 1921 markers and 200 F_{2:3} families, resulting in a total length of 1791.5 cM with an average of 0.9 cM distance between markers. The biggest gap of 21.2 cM was found on chromosome 11 (Table 1 and Figure 2), which may be attributed to lack of polymorphic SNPs or over filtering of variants within this region.

Table 1. Summary of constructed genetic map of F_{2:3} mapping population based on 1921 markers and 200 F_{2:3} families.

Chr.	No. of Markers	Length (cM)	Ave. Interval (cM)	Max Interval (cM)
1	202	199.9	1	9.9
2	193	189.6	1	5.7
3	191	179.4	0.9	11.8
4	128	132.3	1	9.6
5	136	141	1	6.5
6	166	156.3	0.9	9.8
7	161	154.6	1	5.2
8	212	160.2	0.8	6.9
9	116	107.6	0.9	6.2
10	134	97.5	0.7	6.6
11	127	145.1	1.2	21.2
12	155	128.3	0.8	7.4
overall	1921	1791.5	0.9	21.2

2.3.2. QTL Detection

Out of the 300 F_{2:3} families that were phenotyped, only 199 had genotype data. The two datasets were intersected, resulting in a combined dataset of 199 families and 1921 markers used for QTL detection. Logarithm of the odds (LOD) thresholds for 0.05 and 0.01 significance levels were established at 3.92 and 4.76 for interval mapping (IM) and 3.95 and 4.83 for composite interval mapping (CIM) using 10,000 permutations in QGene [42]. Similarly, the LOD thresholds were 3.85 and 4.69 for IM and 3.86 and 4.74 for CIM using QTL Cartographer with 1000 permutations [43]. Based on the analyses, two significant QTLs with LOD scores above the permutation thresholds ($p \leq 0.01$) were identified, and in all cases, the tolerant alleles came from ASD1. These two QTLs, located on the long arm of chromosome 9, *qAG9*, and the short arm of chromosome 7, *qAG7*, were consistently detected by IM and CIM using QGene and QTL Cartographer (Figure 2, Table 2, Figures S2 and S3). The detection by QGene CIM had an LOD score of 15.0 and explains phenotypic variance (R^2) of 29.4 % for *qAG9*, while it reported an LOD score of 7.1 and R^2 of 15.1 % for *qAG7*. Two minor QTLs, with LOD scores slightly below the permutation threshold,

were detected on chromosome 5, *qAG5*, with LOD score of 2.8 and R^2 of 5.4%; and one on chromosome 6, *qAG6*, with LOD score of 2.3 and an R^2 of 5.1%. These two minor QTLs were also derived from ASD1 (Table 2).

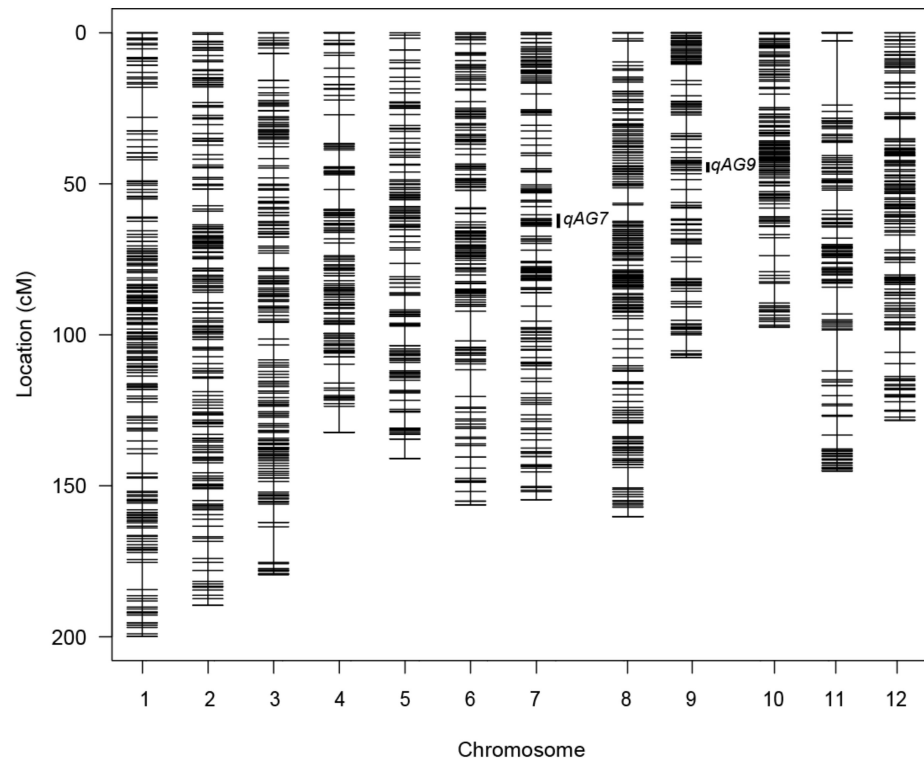


Figure 2. Estimated genetic map of IR64/ASD1 population constructed from 200 $F_{2:3}$ families using 1921 single nucleotide polymorphism (SNP) markers. Each chromosome is represented by a vertical line and each marker by a horizontal line on its respective chromosome position. The positions of the QTLs detected on chromosomes 7 and 9 are represented by black vertical bars.

Table 2. QTLs for anaerobic germination identified in IR64/ASD1 mapping population.

QTL	Chr.	Markers			QGene IM			QGene CIM			QTL Cart IM			QTL Cart CIM		
		Left flank	Peak	Right flank	LOD	R^2	Add	LOD	R^2	Add	LOD	R^2	Add	LOD	R^2	Add
<i>qAG5</i>	5	S05_7886279	S05_9596301	S05_14907902	-	-	-	<i>2.1</i> _a	5.4	2.2	2.1	0.0	0.0	2.9	0.0	-0.5
<i>qAG6</i>	6	S06_10782720	S06_11109090	S06_11200565	2.3	5.1	5.1	2.2	5.0	2.4	2.6	2.7	3.9	2.6	0.3	1.3
<i>qAG7</i>	7	S07_6302957	S07_7547815	S07_7650364	5.6	12.1	8.4	7.1	15.1	8.0	5.5	11.7	8.0	7.5	10.7	8.0
<i>qAG9</i>	9	S09_12029625	S09_12277535	S09_12331208	12.0	24.3	13.0	15.0	29.4	13.0	12.0	30.2	14.0	14.5	27.6	15.0

^a Italicized logarithm of the odds (LOD) scores were below the significant permutation threshold, but with LOD >2

3. Discussion

Four AG QTLs were identified in the QTL analysis, with two above the permutation threshold on chromosomes 7 (*qAG7*) and chromosome 9 (*qAG9*). The QTL with the most statistical support and largest effect, *qAG9*, was detected in the 12.03–12.33 Mb region of chromosome 9 with a peak at 12.28 Mb and is co-located within the *qAG-9-2* region, 11.75–12.55 Mb, as reported by Angaji et al. [2] (Table 3). Kretzschmar et al. [14] have reported, and confirmed by map-based cloning, that a *trehalose-6-phosphate phosphatase* gene, *OsTPP7*, is the causal gene underlying *qAG-9-2*. *OsTPP7* has a locus identifier LOC_Os09g20390 and is reported to be positioned exactly at 12.25 Mb of chromosome 9 according to the rice genome annotation project [44]. Both *qAG9* and *qAG-9-2* were the QTLs with the most statistical support and largest effect detected in their corresponding studies, both having LOD scores greater than the $p \leq 0.01$ significance thresholds and

explaining phenotypic variance of 20–33%. In addition, both studies used IR64 as the sensitive parent in the development of the mapping population. However, unlike in the report of Angaji et al. [2], no QTL was detected from IR64 in this current study.

Table 3. Previously reported QTL for anaerobic germination that overlaps with *qAG9*.

QTL	Source	Marker Position (Mb)			Range (Mb)	Marker	Software	IM		CIM	
		Left Flank	Peak	Right Flank				LOD	R ²	LOD	R ²
<i>qAG-9-2</i>	Khao Hlan On	11.75	12.31	12.55	0.80	136 SSR	QTLCart	20.3	33.5	15.3	20.6
<i>qAG9</i> ^a	ASD1	12.03	12.28	12.33	0.30	RAD-seq	QGene	12.0	24.3	15.0	29.4

^a This QTL was identified in the current study.

The second QTL with the most statistical support and largest effect detected in this study, *qAG7*, was positioned between 6.30–7.65 Mb of chromosome 7, with a peak LOD score at 7.55 Mb. Several other studies have also reported AG tolerance QTL on chromosome 7. A report from Ghosal et al. [4] identified *qSUR7-1_{Rc238-SCR-21}* from the accession Kalarata with a peak at 6.07 Mb (Table 4). There is no information about the flanking region of *qSUR7-1_{Rc238-SCR-21}*, but it seems that the QTL region spans a large interval on chromosome 7 based on the reported graphs. Similarly, the studies of AG tolerance in the reported *qAG7* from Nanhi [12] located at 4.80–18.48 Mb, and in the reported *qAG-7-1* from Khao Hlan On [2] located at 6.07–17.67 Mb span a wide range of QTL region of more than 10 Mb in chromosome 7. Three other QTLs for AG, *qAG7.1* from Ma-Zhan Red [11] located at 8.04 to 12.97 Mb, *qAG7.1* from Kharsu 80A [13] located from 3.68 to 6.01 Mb, and *qAGS-7* from TKM 9 [45] located from 7.61 to 12.78 Mb were reported to be in the similar region but do not overlap entirely with *qAG7*. Among all the studies that reported QTLs for AG in the short arm towards the centromeric region of chromosome 7, this study has shown the narrowest interval of the QTL spans (Table 4).

Table 4. Previously reported QTLs for anaerobic germination that overlap with *qAG7*.

QTL	Source	Marker Position (Mb)			Range (Mb)	Marker	Software	IM		CIM	
		Left Flank	Peak	Right Flank				LOD	R ²	LOD	R ²
<i>qAG7.1</i>	Ma-Zhan Red	8.04	9.07	12.97	4.93	118 SSR, indel	QGene	13.7	30.3	14.5	31.7
<i>qAGS-7</i>	TKM 9	7.61	11.09	12.78	5.17	91 SSR	QGene	7.8	11.1	7.6	17.8
<i>qAG-7-1</i>	Khao Hlan On	6.07	10.37	17.67	11.60	136 SSR	QTLCart	5.1	9.9	-	-
<i>qAG7.1</i>	Kharsu 80A	3.68	4.80	6.01	2.32	BeadXpress	QGene	5.5	17.5	4.8	15.4
<i>qAG7</i>	Nanhi	4.80	NA ^a	18.48	13.68	BeadXpress	QTLCart	11.2	11.7	13.9	14.1
<i>qSUR7-1_{Rc238-SCR-21}</i>	Kalarata	NA	6.07	NA	NA	KASP	IciM	20.3	39.7	NA	NA
<i>qAG7</i> ^b	ASD1	6.30	7.55	7.65	1.35	RAD-seq	QGene	5.6	12.1	7.1	15.1

^a Not applicable (no data available); ^b This QTL was identified in the current study.

The similarities between the results of this study and that of Angaji et al. [2] for *qAG9* and *qAG-9-2* suggests that the same gene, *OsTPP7*, most likely controls their contributions to tolerance to AG, although this needs further confirmation. In contrast, although there have been many reported QTLs in the similar region of *qAG7*, there is not enough evidence currently to say whether *qAG7* and those previously identified QTLs located in the similar region of *qAG7* have the same genetic factor(s) responsible for AG tolerance, due to several non-overlapping regions. Thus, further investigation is needed.

In conclusion, it has been demonstrated in this study that the use of RAD-seq enables the construction of a high-resolution genetic linkage map with an average interval of 0.9 cM. This genotyping method may be used to obtain a precise approximation of QTLs compared to traditional SSR and low-density SNP genotyping methods, which can help us

better understand the genetic basis of AG tolerance in rice. In this study, two significant QTLs for tolerance to anaerobic germination, *qAG7* and *qAG9*, were identified from ASD1. Compared to previous studies, these two QTLs showed more precise locations with the narrowest QTL regions. Though many factors can affect genetic mapping resolution, a higher number of markers can only improve the resolution to a certain extent. Increasing the population size, further backcrossing, or the use of recombinant inbred lines (RILs) can be done to achieve finer mapping resolution. However, these strategies usually require more time and resources. Whether the overlapping QTLs detected in this study and QTLs from other biparental mapping populations in the same chromosomal regions have the same causal gene(s) and/or alleles needs further investigation.

4. Materials and Methods

4.1. Development of Mapping Population

IR64 (GID 4483020), an *indica* mega-variety developed at the International Rice Research Institute (IRRI), which is susceptible to AG [2], was crossed with ASD1 (IRGC 6267; GID 328677), a landrace from India which is tolerant to AG [40,46]. The cross was made in IRRI, Los Baños, Laguna, Philippines in the 2012 wet season. In the 2013 dry season, F₁ individuals were confirmed for true hybrids using SSR markers and were allowed to self-pollinate. In the 2015 dry season, 864 F₂ progenies from a single F₁ individual were harvested to produce F_{2,3} families for phenotyping. The seeds were oven-dried at 40 °C until moisture content was below 14%. The seeds were then stored at 10 °C for 5 months, awaiting phenotyping. Before phenotyping, seeds were incubated in 50 °C for 5 days to break dormancy, and 418 F_{2,3} randomly selected families were tested for germination rate together with parental lines using 40 seeds per entry at 30 °C incubator for 3 days.

4.2. Phenotyping

Measurement of survivability under AG was conducted according to the protocol of Septiningsih et al. [11]. Each tray has 17x34 cells and can accommodate 11 entries with 30 seeds per entry sown at 1 cm depth. An entire column of cells between each entry and the outermost cells were left unseeded to minimize phenotyping error. The *DiGger* R package [47,48] was used to generate 2 replications of alpha plus randomization design with the restriction of having 9 samples and 2 parental controls, ASD1 and IR64, in each tray. Each replicate was laid on an elevated concrete tank filled with tap water up to 10 cm above the tray's surface. The water level was maintained up to 14 days, and then the numbers of seedlings that emerged out of the water were counted for each entry. The average survival of each entry was calculated from the two replicates and was used for the QTL study.

4.3. Genotyping

4.3.1. Library Preparation and Sequencing

Genomic DNA from F₂ leaf samples was isolated using cetyltrimethylammonium bromide (CTAB) extraction method at IRRI, Philippines [49]. DNA was then sent to the Max Plank Institute in Tübingen, Germany, where the DNA libraries were prepared and sequenced. DNA quality was checked with NanoDrop™ 8000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) using 260/230 and 260/280 ratios and quantified using Quant-It™ PicoGreen® dsDNA (Thermo Fisher Scientific, Waltham, MA, USA) quantitation assay. After normalization of DNA concentration, three 96-plex DNA libraries were prepared with *KpnI* restriction enzyme. DNA barcodes for each library were selected from the 386-pool of *KpnI* optimized barcodes to obtain a GC content close to 50% to reduce sequencing errors. Barcoded DNA fragments for each library were multiplexed in a microcentrifuge tube and sheared with Covaris S220 Focused-ultrasonicator (Covaris, Woburn, MA, USA). End-repair and dA-tailing were performed after DNA shearing; then, universal sequencing adapters were ligated. After a PCR enrichment step, libraries were selected for 300–600 bp fragment size using BluePippin (Sage Science, Beverly, MA, USA)

automated gel extraction. Libraries were validated for concentration and size distribution with Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) using a DNA 1000 chip. The resulting libraries were sequenced with Illumina HiSeq 2000 (Illumina, San Diego, CA, USA), each on separate lanes with 100-bp single-end reads.

4.3.2. Variant Detection

PCR and optical duplicate reads were removed prior to sequence processing with BBMap using the *clumpify* function [50]. The duplicated reads had the same sequence at both 5' and 3' ends, which were produced by either duplication from the PCR amplification step during the library preparation or during the sequencing process. The reads of similar sequence were sorted together in the process, allowing better compression of the remaining reads. This also allows faster sequence processing and more accurate variant calling. Sequence demultiplexing was done using GBSX software [51] identifying barcode indices with 0 mismatches followed by *KpnI* cut sites, producing individual FastQ files for each sample and clipping out the barcode sequences. FastQ files of each parental line from three lanes were concatenated together. Trimmomatic [52] was used to filter reads shorter than 36 bp, trim the start or end of reads having quality lower than 5 (>0.31 probability error), and clipping Illumina adapter sequences by allowing 2 mismatches in the initial 16 bp seed sequence and entirely clipping the adapter when a matching score reaches 10 (about 17 bp match). The resulting sequences were aligned to rice reference genome [53], *Oryza sativa* L. ssp. *japonica* cv. Nipponbare, using the sequence aligner, Burrows-Wheeler Alignment maximal exact matches [54]. The aligned reads were sorted and assigned groups using Picard [55]. Genome Analysis Toolkit (GATK) *BaseRecalibrator* [56] was performed on the alignment files using initially called homozygous variant sites between the two parental lines as known sites. Genomic variants for each sample were identified using GATK *HaplotypeCaller* and stored into an instance of GenomicsDB datastore [57]. Variants for all samples were extracted from the datastore into VCF files using GATK *GenotypeVCFs* command.

4.3.3. Variant and Sample Filtering

Variant data from restriction enzyme-based genotyping was filtered before adapting genotype data to F₂ population genetic analysis by following some steps from Furuta et al. [41]. To simplify further analysis, only SNPs with exactly 2 allele counts were selected using BCFtools [58]. Genotype calls with less than 10 allele depth (read count) or less than 30-phred genotyping quality were set to missing calls using VCFtools [59].

SNPs were further filtered by selecting homozygous and polymorphic SNPs between the two parents using Trait Analysis by aSSociation, Evolution and Linkage (TASSEL) [60]. SNPs that have less than 0.18 minimum allele frequencies, and heterozygous proportions less than 0.05 and greater than 0.55 were excluded from the dataset. Redundant SNPs that were found on the same read, within 96 bp from one another, were removed using the *thin* parameter of VCFtools. Samples with at least call rate of 50% and markers with call rate of at least 60% were selected using TASSEL and were used for imputation and genotyping call correction.

4.3.4. Adapting genotyping Data for F₂ Genetics Analysis

Prior to QTL mapping, adapting to genotyping data for sequencing-based genotyping was done as described by Furuta et al. [41]. "ABH Genotype" file was generated using TASSEL by setting IR64 as parent A and ASD1 as parent B. The genotype file was loaded into RStudio [61] and adjusted using the *ABHgenotypeR* R package [47,62]. Missing data was imputed then corrected for heterozygous and homozygous calls by 10 and 7 bp lengths, respectively. Two additional steps were done further to reduce potentially erroneous calls that were left out by the previous steps. The first was to set remaining double recombinants to missing (e.g., in the sequence HHAABB, AA were replaced by missing data, NN), and the second was to set erroneous calls at the start and end of chromosomes

to missing (e.g., in the chromosome start sequence of 5'-AAHH or end sequence BBAA-3', both AAs were replaced by NN). These two steps were done by correcting all observed patterns with ≤ 4 bp in length using the *gtmask* R package that we have written (<https://github.com/johncarlosignacio/gtmask>, accessed on 3 April 2021).

4.4. QTL Analysis

4.4.1. Genetic Map Estimation

The output from *ABHgenotypeR* was loaded as cross data of *R/qtl* package [63]. Duplicate markers with the same non-missing genotype calls and adjacent were removed so that only one marker with the same genotype remains. This was done by using the *R/qtl findDupMarkers* function with the parameters *exact.only* set to false and *adjacent.only* set to true. The genetic map was estimated using the *R/qtl est.map* function with Kosambi as mapping function.

4.4.2. QTL Detection Using QGene Software

The final mapping dataset was obtained by intersecting the $F_{2,3}$ families that had both phenotypic and genotypic data. The intersection between these two datasets was used for QTL analysis. Interval mapping (IM) and composite interval mapping (CIM) analyses were first done using QGene [42]. QTL scan interval was set to 1.0 cM for both analyses. For CIM, cofactors were selected using stepwise cofactor selection as the method with maximum number of cofactors set to auto, F to add set to 0.01, and F to drop set to 0.01. Thresholds for 0.05 and 0.01 significances were obtained using 10,000 permutations.

4.4.3. QTL Selection Using QTL Cartographer Software

The data was also analyzed using WinQTL Cartographer for comparison [43]. For both IM and CIM analyses, walk speed was set to 1.0 cM. For CIM, the model used was no. 6 (standard model), automatically selecting 5 control markers, setting window size to 1.0 cM, and forward & backward regression probabilities for into and out of 0.01. Thresholds for 0.05 and 0.01 significances were obtained using 1000 permutations.

4.5. QTL Comparisons

Detected QTLs in this study were compared to previously published QTLs to determine whether the detected QTLs were novel. If previous studies presented genetic map information, the map positions were converted into physical positions by querying the corresponding primer sequences against the rice reference genome [44] with nucleotide basic local alignment search tool (BLAST) [64].

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/plants10040705/s1>, Figure S1: Representative tray showing different levels of survival rates, Figure S2: Plot of LOD from interval mapping QTL analysis of $F_{2,3}$ mapping population derived from IR64 and ASD1 on chromosomes 5, 6, 7, and 9 from WinQTL Cartographer, Figure S3: Plot of LOD from composite interval mapping QTL analysis of $F_{2,3}$ mapping population derived from IR64 and ASD1 on chromosomes 5, 6, 7, and 9 from WinQTL Cartographer, Table S1: Distribution of genotypes at each data correction step.

Author Contributions: Conceptualization, E.M.S. and D.W.; methodology, J.C.I.I., M.Z., D.W., and E.M.S.; software, J.C.I.I.; validation, J.S.M., M.S.M., and E.M.S.; formal analysis, J.C.I.I.; investigation, J.C.I.I., C.C.J., and E.M.S.; data curation, J.C.I.I.; writing—original draft preparation, J.C.I.I.; writing—review and editing, J.C.I.I., J.M.S., M.S.M., S.D., T.K., D.W., and E.M.S.; visualization, J.C.I.I.; supervision, E.M.S. and T.K.; project administration, J.C.I.I., M.Z., S.D., D.W., and E.M.S.; funding acquisition, E.M.S. and D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Federal Ministry for Economic Cooperation and Development (BMZ) grant number #81157485 to E.M.S. and D.W.

Data Availability Statement: The genotype data of the mapping population is accessible in the *gtmask* R package (<https://github.com/johncarlosignacio/gtmask>, accessed on 3 April 2021).

Acknowledgments: Josie Mendoza, Eloise Suiton, Irma Tamisin, Dessa Gutierrez, Liza Bugtong, Neil Bantasan, Freddie Perez, Jojo Borgonia, and Reynaldo Garcia for technical support.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Pandey, S.; Mortimer, M.; Wade, L.; Tuong, T.; Lopez, K.; Hardy, B. Direct seeding: Research issues and opportunities. In Proceedings of the International Workshop on Direct Seeding in Asian Rice Systems: Strategic Research Issues and Opportunities, Bangkok, Thailand, 25–28 January 2000; International Rice Research Institute: Los Baños, Philippines, 2002; pp. 3–12.
- Angaji, S.A.; Septiningsih, E.M.; Mackill, D.J.; Ismail, A.M. QTLs associated with tolerance of flooding during germination in rice (*Oryza sativa* L.). *Euphytica* **2010**, *172*, 159–168. [[CrossRef](#)]
- Ismail, A.M.; Ella, E.S.; Vergara, G.V.; Mackill, D.J. Mechanisms associated with tolerance to flooding during germination and early seedling growth in rice (*Oryza sativa*). *Ann. Bot.* **2009**, *103*, 197–209. [[CrossRef](#)]
- Ghosal, S.; Casal, C.; Quilloy, F.A.; Septiningsih, E.M.; Mendiolo, M.S.; Dixit, S. Deciphering Genetics Underlying Stable Anaerobic Germination in Rice: Phenotyping, QTL Identification, and Interaction Analysis. *Rice* **2019**, *12*. [[CrossRef](#)]
- Alam, R.; Hummel, M.; Yeung, E.; Locke, A.M.; Ignacio, J.C.I.; Baltazar, M.D.; Jia, Z.; Ismail, A.M.; Septiningsih, E.M.; Bailey-Serres, J. Flood resilience loci *SUBMERGENCE 1* and *ANAEROBIC GERMINATION 1* interact in seedlings established underwater. *Plant Direct* **2020**, *4*, e00240. [[CrossRef](#)] [[PubMed](#)]
- Mondal, S.; Khan, M.I.R.; Entila, F.; Dixit, S.; Pompe, P.C.; Panna Ali, M.; Pittendrigh, B.; Septiningsih, E.M.; Ismail, A.M. Responses of AG1 and AG2 QTL introgression lines and seed pre-treatment on growth and physiological processes during anaerobic germination of rice under flooding. *Sci. Rep.* **2020**, *10*, 10214. [[CrossRef](#)] [[PubMed](#)]
- Mondal, S.; Khan, M.I.R.; Dixit, S.; Pompe, P.C.; Septiningsih, E.M.; Ismail, A.M. Growth, productivity and grain quality of AG1 and AG2 QTLs introgression lines under flooding in direct-seeded rice system. *Field Crop. Res.* **2020**, *248*, 107713. [[CrossRef](#)]
- Toledo, A.M.U.; Ignacio, J.C.I.; Casal, C.; Gonzaga, Z.J.; Mendiolo, M.S.; Septiningsih, E.M. Development of Improved Ciherang-Sub1 Having Tolerance to Anaerobic Germination Conditions. *Plant Breed. Biotechnol.* **2015**, *3*, 77–87. [[CrossRef](#)]
- Ghosal, S.; Quilloy, F.A.; Casal, C.; Septiningsih, E.M.; Mendiolo, M.S.; Dixit, S. Trait-based mapping to identify the genetic factors underlying anaerobic germination of rice: Phenotyping, GXE, and QTL mapping. *BMC Genet.* **2020**, *21*, 6. [[CrossRef](#)]
- Miro, B.; Ismail, A.M. Tolerance of anaerobic conditions caused by flooding during germination and early growth in rice (*Oryza sativa* L.). *Front. Plant Sci.* **2013**, *4*, 269. [[CrossRef](#)]
- Septiningsih, E.M.; Ignacio, J.C.I.; Sendon, P.M.D.; Sanchez, D.L.; Ismail, A.M.; Mackill, D.J. QTL mapping and confirmation for tolerance of anaerobic conditions during germination derived from the rice landrace Ma-Zhan Red. *Theor. Appl. Genet.* **2013**, *126*, 1357–1366. [[CrossRef](#)]
- Baltazar, M.D.; Ignacio, J.C.I.; Thomson, M.J.; Ismail, A.M.; Mendiolo, M.S.; Septiningsih, E.M. QTL mapping for tolerance of anaerobic germination from IR64 and the aus landrace Nanhi using SNP genotyping. *Euphytica* **2014**, *197*, 251–260. [[CrossRef](#)]
- Baltazar, M.D.; Ignacio, J.C.I.; Thomson, M.J.; Ismail, A.M.; Mendiolo, M.S.; Septiningsih, E.M. QTL mapping for tolerance to anaerobic germination in rice from IR64 and the aus landrace Kharsu 80A. *Breed. Sci.* **2019**, *69*, 227–233. [[CrossRef](#)] [[PubMed](#)]
- Kretzschmar, T.; Pelayo, M.A.F.; Trijatmiko, K.R.; Gabunada, L.F.M.; Alam, R.; Jimenez, R.; Mendiolo, M.S.; Slamet-Loedin, I.H.; Sreenivasulu, N.; Bailey-Serres, J.; et al. A trehalose-6-phosphate phosphatase enhances anaerobic germination tolerance in rice. *Nat. Plants* **2015**, *1*. [[CrossRef](#)] [[PubMed](#)]
- Chamara, B.S.; Marambe, B.; Kumar, V.; Ismail, A.M.; Septiningsih, E.M.; Chauhan, B.S. Optimizing Sowing and Flooding Depth for Anaerobic Germination-Tolerant Genotypes to Enhance Crop Establishment, Early Growth, and Weed Management in Dry-Seeded Rice (*Oryza sativa* L.). *Front. Plant Sci.* **2018**, *9*, 1654. [[CrossRef](#)]
- Lal, B.; Gautam, P.; Nayak, A.K.; Raja, R.; Shahid, M.; Tripathi, R.; Singh, S.; Septiningsih, E.M.; Ismail, A.M. Agronomic manipulations can enhance the productivity of anaerobic tolerant rice sown in flooded soils in rainfed areas. *Field Crop. Res.* **2018**, *220*, 105–116. [[CrossRef](#)]
- Septiningsih, E.M.; Mackill, D.J. Genetics and breeding of flooding tolerance in rice. In *Rice Genomics, Genetics and Breeding*; Springer: Singapore, 2018; pp. 275–295. ISBN 9789811074615.
- Ling, J.; MingYu, H.; ChunMing, W.; JianMin, W. Quantitative trait loci and epistatic analysis of seed anoxia germinability in rice (*Oryza sativa*). *Rice Sci.* **2004**, *11*, 238–244.
- Baird, N.A.; Etter, P.D.; Atwood, T.S.; Currey, M.C.; Shiver, A.L.; Lewis, Z.A.; Selker, E.U.; Cresko, W.A.; Johnson, E.A. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **2008**, *3*. [[CrossRef](#)]
- 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](#)]
- Poland, J.A.; Rife, T.W. Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome* **2012**, *5*, 92–102. [[CrossRef](#)]
- Zhou, L.; Wang, S.B.; Jian, J.; Geng, Q.C.; Wen, J.; Song, Q.; Wu, Z.; Li, G.J.; Liu, Y.Q.; Dunwell, J.M.; et al. Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Sci. Rep.* **2015**, *5*, 9350. [[CrossRef](#)] [[PubMed](#)]

23. Liu, N.; Li, M.; Hu, X.; Ma, Q.; Mu, Y.; Tan, Z.; Xia, Q.; Zhang, G.; Nian, H. Construction of high-density genetic map and QTL mapping of yield-related and two quality traits in soybean RILs population by RAD-sequencing. *BMC Genom.* **2017**, *18*, 466. [CrossRef] [PubMed]
24. Wang, X.; Cheng, Y.; Yang, C.; Yang, C.; Mu, Y.; Xia, Q.; Ma, Q. QTL mapping for aluminum tolerance in RIL population of soybean (*Glycine max* L.) by RAD sequencing. *PLoS ONE* **2019**, *14*, e0223674. [CrossRef]
25. Wang, L.; Conteh, B.; Fang, L.; Xia, Q.; Nian, H. QTL mapping for soybean (*Glycine max* L.) leaf chlorophyll-content traits in a genotyped RIL population by using RAD-seq based high-density linkage map. *BMC Genom.* **2020**, *21*, 739. [CrossRef] [PubMed]
26. Liang, Y.; Baring, M.; Wang, S.; Septiningsih, E.M. Mapping QTLs for Leafspot Resistance in Peanut Using SNP-Based Next-Generation Sequencing Markers. *Plant Breed. Biotechnol.* **2017**, *5*, 115–122. [CrossRef]
27. Liang, Y.; Baring, M.R.; Septiningsih, E.M. Mapping of quantitative trait loci for yield and grade related traits in peanut (*Arachis hypogaea* L.) using high-resolution SNP markers. *Plant Breed. Biotechnol.* **2018**, *6*, 454–462. [CrossRef]
28. Liang, Y.; Cason, J.M.; Baring, M.R.; Septiningsih, E.M. Identification of QTLs associated with Sclerotinia blight resistance in peanut (*Arachis hypogaea* L.). *Genet. Resour. Crop Evol.* **2020**, *68*, 629–637. [CrossRef]
29. Pan, L.; Wang, N.; Wu, Z.; Guo, R.; Yu, X.; Zheng, Y.; Xia, Q.; Gui, S.; Chen, C. A High Density Genetic Map Derived from RAD Sequencing and Its Application in QTL Analysis of Yield-Related Traits in *Vigna unguiculata*. *Front. Plant Sci.* **2017**, *8*, 1544. [CrossRef] [PubMed]
30. Matsumura, H.; Miyagi, N.; Taniai, N.; Fukushima, M.; Tarora, K.; Shudo, A.; Urasaki, N. Mapping of the Gynoecy in Bitter Gourd (*Momordica charantia*) Using RAD-Seq Analysis. *PLoS ONE* **2014**, *9*, e87138. [CrossRef] [PubMed]
31. Cui, J.; Luo, S.; Niu, Y.; Huang, R.; Wen, Q.; Su, J.; Miao, N.; He, W.; Dong, Z.; Cheng, J.; et al. A RAD-Based Genetic Map for Anchoring Scaffold Sequences and Identifying QTLs in Bitter Gourd (*Momordica charantia*). *Front. Plant Sci.* **2018**, *9*, 477. [CrossRef] [PubMed]
32. Laila, R.; Park, J.I.; Robin, A.H.K.; Natarajan, S.; Vijayakumar, H.; Shirasawa, K.; Isobe, S.; Kim, H.T.; Nou, I.S. Mapping of a novel clubroot resistance QTL using ddRAD-seq in Chinese cabbage (*Brassica rapa* L.). *BMC Plant Biol.* **2019**, *19*, 13. [CrossRef]
33. Zhang, F.; Kang, J.; Long, R.; Yu, L.X.; Wang, Z.; Zhao, Z.; Zhang, T.; Yang, Q. High-density linkage map construction and mapping QTL for yield and yield components in autotetraploid alfalfa using RAD-seq. *BMC Plant Biol.* **2019**, *19*, 165. [CrossRef]
34. Wu, K.; Liu, H.; Yang, M.; Tao, Y.; Ma, H.; Wu, W.; Zuo, Y.; Zhao, Y. High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (*Sesamum indicum* L.) based on RAD-Seq technology. *BMC Plant Biol.* **2014**, *14*, 274. [CrossRef]
35. Wang, J.; Wang, Z.; Du, X.; Yang, H.; Han, F.; Han, Y.; Yuan, F.; Zhang, L.; Peng, S.; Guo, E. A high-density genetic map and QTL analysis of agronomic traits in foxtail millet [*Setaria italica* (L.) P. Beauv.] using RAD-seq. *PLoS ONE* **2017**, *12*, e0179717. [CrossRef] [PubMed]
36. Kajiya-Kanegae, H.; Takanashi, H.; Fujimoto, M.; Ishimori, M.; Ohnishi, N.; Wacera, W. F.; Omollo, E.A.; Kobayashi, M.; Yano, K.; Nakano, M.; et al. RAD-seq-Based High-Density Linkage Map Construction and QTL Mapping of Biomass-Related Traits in Sorghum using the Japanese Landrace Takakibi NOG. *Plant Cell Physiol.* **2020**, *61*, 1262–1272. [CrossRef] [PubMed]
37. Peng, Y.; Hu, Y.; Mao, B.; Xiang, H.; Shao, Y.; Pan, Y.; Sheng, X.; Li, Y.; Ni, X.; Xia, Y.; et al. Genetic analysis for rice grain quality traits in the YVB stable variant line using RAD-seq. *Mol. Genet. Genom.* **2016**, *291*, 297–307. [CrossRef] [PubMed]
38. Hsieh, M.; Wu, D.; Chen, K. Mapping of QTLs controlling pre-harvest sprouting in the cross between japonica rice cultivars using RAD sequencing. *Crop. Environ. Bioinform.* **2014**, *11*, 11–25.
39. Rebolledo, M.C.; Dingkuhn, M.; Clément-Vidal, A.; Rouan, L.; Luquet, D. Phenomics of rice early vigour and drought response: Are sugar related and morphogenetic traits relevant? *Rice* **2012**, *5*, 22. [CrossRef] [PubMed]
40. Yamauchi, M.; Winn, T. Rice seed vigor and seedling establishment in anaerobic soil. *Crop Sci.* **1996**, *36*, 680–686. [CrossRef]
41. Furuta, T.; Ashikari, M.; Jena, K.K.; Doi, K.; Reuscher, S. Adapting genotyping-by-sequencing for rice F2 populations. *G3 Genes Genomes Genet.* **2017**, *7*, 881–893. [CrossRef] [PubMed]
42. Nelson, J.C. QGENE: Software for marker-based genomic analysis and breeding. *Mol. Breed.* **1997**, *3*, 239–245. [CrossRef]
43. Wang, S.; Basten, C.J.; Zeng, Z.-B. *Windows QTL Cartographer 2.5*; Department of Statistics, North Carolina State University: Raleigh, NC, USA, 2012; Available online: <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm> (accessed on 12 December 2019).
44. Kawahara, Y.; de la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Wu, J.; Zhou, S.; et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [CrossRef]
45. Myint, O.; Hernandez, J.E.; Ismail, A.M.; Septiningsih, E.M. Mapping QTLs for tolerance to flooding during germination and early seedling growth in rice (*Oryza sativa* L.) using IR64 × TKM9 population. In Proceedings of the Seventh Agricultural Research Conference, Nay Pyi Taw, Myanmar, 10–11 January 2013; pp. 81–91.
46. Biswas, J.K.; Yamauchi, M. Mechanism of seedling establishment of direct-seeded rice (*Oryza sativa* L.) under lowland conditions. *Bot. Bull. Acad. Sin.* **1997**, *38*, 29–32.
47. Developer Core Team, R. R: *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
48. Coombes, N.E. *DiGger: DiGger Design Generator under Correlation and Blocking*; NSW DPI Biometrics: Sydney, Australia, 2011.
49. Doyle, J.; Doyle, J.L. Genomic plant DNA preparation from fresh tissue—CTAB method. *Phytochem. Bull.* **1987**, *19*, 11–15.
50. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*; LBNL Department of Energy Joint Genome Institute: Walnut Creek, CA, USA, 2014.

51. Herten, K.; Hestand, M.S.; Vermeesch, J.R.; Van Houdt, J.K.J. GBSX: A toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinf.* **2015**, *16*. [[CrossRef](#)] [[PubMed](#)]
52. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
53. Matsumoto, T.; Wu, J.; Kanamori, H.; Katayose, Y.; Fujisawa, M.; Namiki, N.; Mizuno, H.; Yamamoto, K.; Antonio, B.A.; Baba, T.; et al. The map-based sequence of the rice genome. *Nature* **2005**, *436*, 793–800. [[CrossRef](#)]
54. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
55. Broad Institute. Picard Toolkit. GitHub Repository. 2019. Available online: <http://broadinstitute.github.io/picard/> (accessed on 12 December 2019).
56. Poplin, R.; Ruano-Rubio, V.; DePristo, M.A.; Fennell, T.J.; Carneiro, M.O.; Van der Auwera, G.A.; Kling, D.E.; Gauthier, L.D.; Levy-Moonshine, A.; Roazen, D.; et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* **2017**, 201178. [[CrossRef](#)]
57. Datta, K.; Gururaj, K.; Naik, M.; Narvaez, P.; Rutar, M. “GenomicsDB - Sparse Array Storage for Genomics.” 2019. Available online: <https://www.genomicsdb.org/> (accessed on 12 December 2019).
58. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [[CrossRef](#)] [[PubMed](#)]
59. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)] [[PubMed](#)]
60. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)]
61. RStudio Team. *RStudio: Integrated Development Environment for R*; RStudio PBC: Boston, MA, USA, 2019.
62. Reuscher, S.; Furuta, T. ABHgenotypeR: Easy Visualization of ABH Genotypes. Available online: <https://CRAN.R-project.org/package=ABHgenotypeR> (accessed on 12 December 2019).
63. Broman, K.W.; Wu, H.; Sen, S.; Churchill, G.A. R/qtl: {QTL} mapping in experimental crosses. *Bioinformatics* **2003**, *19*, 889–890. [[CrossRef](#)] [[PubMed](#)]
64. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]