

RESEARCH ARTICLE

Assessment of the association between yield parameters and polymorphic sites of the *Ghd7* locus in a core-set of rice cultivars grown in Sri Lanka

N.D.U.S. Nakandala¹, S.I. Karunarathne¹, L.T. Ranaweera¹, W.W.M.U.K. Wijesundara¹,
R.W.K.M. Senevirathna¹, R.M.S.K. Rathnayake¹, U.A.K.S. Udawela² and S.D.S.S. Sooriyapathirana^{1,*}

¹Department of Molecular Biology and Biotechnology, Faculty of Science, University of Peradeniya, 20400, Peradeniya, Sri Lanka

²Rice Research and Development Institute, Bathalagoda, Ibbagamuwa, 60500, Sri Lanka

Received: 21/08/2018 ; Accepted: 01/01/2019

Abstract: Yield improvement is a major aspect in rice breeding programs. *Ghd7* is a pleiotropic gene which regulates yield, plant height, and heading date in rice. Although *Ghd7* has been previously cloned and sequenced in several other rice cultivars, none of the studies have been conducted for Sri Lankan rice germplasm. Therefore, in this study, we assessed the yield performance of 12 rice cultivars, genetic polymorphism of *Ghd7* locus and associations between DNA markers and yield traits. Breeder seeds of the selected cultivars were obtained from RRDI, Bathalagoda, and established under greenhouse conditions at Peradeniya, Sri Lanka in *Maha* and *Yala* seasons, 2017. The cultivars were subjected to morphometric analysis, genotyped for 12 DNA markers and sequenced at *Seq7-8* locus. Morphometric data were subjected to single marker analysis using General Linear Model (GLM) in SAS 9.4. Here we reported six marker haplotypes based on the arrangement of 13 DNA marker alleles at *Ghd7*. Moreover, twenty-three SNP/INDEL variations at *Seq7-8* locus revealed close genetic relationships between the rice cultivars Bg 90-2, Bg 352 and At 307, Bg310. Four polymorphic markers (*Seq7-8*, *Seq1-2*, *RM5436* and *RM5346*) were significantly associated with rice yield traits so that they could be used in marker assisted selection. The SNPs/INDELS of *Ghd7* were significantly associated with all the yield traits except 100 seed weight and 100 endosperm weight. Thus the present study demonstrates the possibility of employing marker assisted breeding to improve rice yield using the polymorphic genomic information at *Ghd7* locus.

Keywords: DNA markers for rice, Rice marker haplotypes, Pleiotropic genes in rice, Rice breeding in Sri Lanka, SNP and INDEL in rice breeding.

INTRODUCTION

Rice yield is a complex agronomic trait governed by both the genetic constitution of the plant and the environment (Yan *et al.*, 2011). Moreover, rice yield is a combination of three individual traits; namely the grain yield, plant height and heading date (Xue *et al.*, 2008). The grain yield in turn depends on the number of panicles per plant, number of grains per panicle and grain weight (Huang *et al.*, 2009). Plant height regulates the plant architecture while the

heading date facilitates the environmental adaptation of rice. Also, the plant height and the heading date collectively have a greater contribution on rice yield (Xue *et al.*, 2008).

Genetically, rice yield, governed by quantitative trait loci (QTLs), shows continuous variation in segregating populations. Thus individual QTLs have minor effects on overall variations of the yield (Yano and Sasaki, 1997). *Ghd7* has been identified as one of such important QTLs, which plays a key role in improving the rice yield by simultaneous regulation of the plant height, grain number and the heading date (Xue *et al.*, 2008). Higher expression of *Ghd7* causes increased number of tillers, taller plants and delayed heading dates. *Ghd7* is a 3,917bp long gene and its cDNA is 1,014bp long. The two exon regions in *Ghd7* encode a CCT domain protein which contains 257 amino acids (Xing *et al.*, 2014; Xue *et al.*, 2008). This protein was identified in both *indica* rice varieties such as *Minghui 63* and *japonica* rice varieties such as *Nipponbare* (Xue *et al.*, 2008; Lu *et al.*, 2012).

Previous studies have revealed the features of *Ghd7* in many rice germplasms belonging to *indica* and *japonica* varieties. Lu *et al.*, (2012) studied the nucleotide diversity of the whole *Ghd7* gene among a set of *Oryza sativa* cultivars. Studies have also shown that *Ghd7* mediated flowering can be observed not only in rice, but also in other grasses such as maize, sorghum and purple false brome (*Brachypodium distachyon*) (Yang *et al.*, 2012). Xing *et al.*, (2014) has successfully cloned and mapped *Ghd7* to a position of 0.4 cM in Chromosome 7 in rice using *Minghui 63* and *Zhenshan 97* cultivars. The haplotypes based on SNPs and protein diversity of *Ghd7* also have been identified (Xue *et al.*, 2008; Lu *et al.*, 2012). Moreover, the regulation of *Ghd7* expression by various environmental signals (Weng *et al.*, 2014) and its effects on other genes in flowering pathway of rice have been studied in detail (Matsubara *et al.*, 2011). Furthermore, the involvement of *Ghd7* in regulating various other aspects of plant development such as plant architecture, branching of tillers, hormone metabolism and plant responses to various environmental stresses have been reported (Weng *et al.*, 2014).

*Corresponding Author's Email: sunethuop@gmail.com

 <http://orcid.org/0000-0002-5592-1742>



However, the previous studies on *Ghd7* have not considered the Sri Lankan rice germplasm. Therefore, the present study was focused on 10 Sri Lankan rice varieties and two landraces (hereafter mentioned as cultivars) in order to assess the potential of using the *Ghd7* polymorphism in marker assisted breeding (MAB). The selected set of cultivars represents three yield-classes (high, moderate and low). We aimed to detect the genetic polymorphism of the *Ghd7* locus among rice cultivars using DNA markers linked to the *Ghd7* locus and then to assess the associations of the yield traits with DNA markers and SNP/INDEL variations present within the *Seq7-8* locus which is located within *Ghd7*.

MATERIALS AND METHODS

Morphometric analysis

Plant material

A total of 12 rice cultivars under three main yield categories; high, moderate and low were selected and breeder seeds were obtained for planting from the Rice Research and Development Institute (RRDI) Bathalagoda, Sri Lanka (Table 1).

Table 1: Important characteristics and recommended cultivation areas of selected cultivars for the study

Yield Category (t/ha)	Cultivar/ Yield (t/ha)	Type of cultivar*	Cultivation area	Important characteristics
High yielding (5.5-6.5)	At362 (5.5-6.5)	NI	LLWZ	Red pericarp, moderately resistant to brown plant hopper, blast and bacterial blight, highest tolerance for Phosphorus deficiency, moderately Fe tolerant (Priyantha <i>et al.</i> , 2013)
	Bg366 (5.0-6.0)	NI	GC	White pericarp, non-glutinous endosperm, intermediate bold grain type, resistant /moderate resistant to blast, resistant to bacterial leaf blight, moderate resistant to gall midge and brown plant hopper (Bentota, 2009)
	Bg90-2 (5.5-6.5)	NI	GC	High yielding (RRDI, 2017)
Moderate yielding (4.5-5.5)	Bg300 (4.5-5.5)	NI	GC	Early maturity, high adaptability, resistant to gall midge (Biotype1), brown plant hopper, bacterial blight, blast, moderately Fe ²⁺ tolerant (Priyantha <i>et al.</i> , 2013; DOA, 2017; RRDI, 2017)
	At307 (4.5-5.5)	NI	GC	White pericarp, resistant to brown plant hopper, gall midge and blast, moderately Fe ³⁺ tolerant (Priyantha <i>et al.</i> , 2013)
	Bg310 (4.5-5.5)	NI	-	White pericarp, intermediate bold grain type, resistant or moderately resistant to gall midge and brown plant hopper, blast, salt tolerant (DOA, 2015; Hemachandra and Amarasinghe, 2016)
	Bg352 (5.0-5.5)	NI	GC	White pericarp, intermediate bold type grain, earliness, wide adaptability, moderately tolerant for Phosphorus deficiency, resistant to blast, gall midge (Biotype1), moderately resistant to bacterial leaf blight and brown plant hopper, sensitive to Fe ²⁺ toxicity, moderately sensitive to thrips (RRDI, 2017; DOA, 2017)
Low yielding (2.5-4.5)	Bw367 (5.0-5.5)	NI	LL	White pericarp, highly Fe ²⁺ tolerant (Priyantha <i>et al.</i> , 2013)
	Bg250 (3.5-4.5)	NI	GC	White pericarp, ultra-short in maturity, resistant/moderately resistant to brown plant hopper and blast, highly tolerant to salt, field level of Fe ²⁺ tolerance (Priyantha <i>et al.</i> , 2013)
	Bw272-6b (3.5-4.5)	NI	LLWZ	Red pericarp, resistant to blast, resistant to lodging, highly Fe ²⁺ susceptible (DOA, 2017; Priyantha <i>et al.</i> , 2013; RRDI, 2017)
	<i>Pachchaperumal</i> (2.5-3.0)	T	GC	Red pericarp, medium grain shape, help control diabetes and cardiovascular diseases, nutritional value (IUCN, 2016)
	<i>Suwadhah</i> (2.5-2.75)	T	GC	White pericarp, exquisite taste, fragrant aroma, promote fair and glowing skin, improve the function of excretory system, improve vocal clarity, help control diabetes (IUCN, 2016; DOA, 2017)

*NI indicates newly improved genotypes; *T indicates the traditional cultivars in Sri Lanka; GC indicates general cultivation, LLWZ: lowland wet zone

Planting and sample collection for DNA extraction

Rice seeds were soaked in water for 24 hours and wrapped in a wet tissue paper until germination. About three quarter of each pot was filled with soil and water (~ 2.5 cm) was added over the soil surface. The plants were managed in a greenhouse at University of Peradeniya, Sri Lanka according to the standard agronomic practices recommended by the Department of Agriculture (DOA), Sri Lanka. The pots were laid out as a Completely Randomized Design (CRD). Ten seeds were placed on each pot and four pots were maintained per each cultivar. In each pot, four plants were selected after two-weeks based on their health and growth and the rest of the plants were discarded. The trial was repeated in both *Maha* and *Yala* seasons in the year 2017. Leaf samples were collected from each cultivar at the age of three weeks. The collected samples were crushed in liquid nitrogen and stored at -80 °C.

Morphological data for marker-trait association analyses

The vegetative growth parameters; plant height (PIH), culm length (CL), number of tillers (NT), leaf blade length (LBL) and leaf blade width (LBW) were measured in 16 plants per each cultivar at the ages of 3, 6, 9, 12 and 15 weeks for both *Maha* and *Yala* seasons. Then, the reproductive parameters; heading date (HD) (days for the first panicle to emerge), flag leaf length (FLL), flag leaf width (FLW) were recorded at blooming. Simultaneously, the harvesting parameters; the days to harvest from the initial establishment, overall yield per plant, seed number per plant (SN), seed length (SL), seed width, seed weight, 100 seed weight (HSW) and 100 endosperm weight (HEW), endosperm length (EL) and endosperm width (EW) were measured. All the parameters were assessed with reference to the previous reports on morphometric measurements in rice (Xue et al., 2008; Zhang et al., 2015; Liu et al., 2013).

Genetic analysis*DNA extraction, PCR and sequencing*

The genomic DNA was extracted from the crushed immature leaf samples using CTAB method (Nageswara-Rao et al., 2013). Then, DNA was amplified using seven DNA markers (Table 2). The PCR mixture contained the total volume of 10 µl which comprised of 5 µl of 2×Go Taq Green Master Mix (Promega Madison, WI USA), 0.5 µl each forward and reverse primers (10 µM), 3 µl nuclease free water and 1µl of respective template DNA (50 - 80 ng/µl). The amplification was performed in a Thermal Cycler (TP600: Takara, Otsu Shiga, Japan) under the conditions; initial denaturation at 94 °C for 5 mins, followed by 35 cycles of 30 sec denaturation at 94 °C, 1 min annealing at relevant primer annealing temperatures (Table 2), and 2 mins extension at 72 °C followed by a final extension at 72 °C for 10 mins. The amplified PCR products were resolved using 1.5% agarose gel electrophoresis. The *Seq7-8* PCR products were purified using the QIA quick PCR purification kit (Catalog No: 28104, Qiagen, Hilden, Germany) and then sanger-sequenced using the Genetic Analyzer 3500 (Catalog No:622-0010, Applied Bio System®). All the generated sequences are found under the Genbank accession numbers; MH779562-MH779573.

Data analysis*DNA marker length polymorphism of Ghd7 and association with yield traits*

All the collected morphometric data were subjected to normality test. Then the marker-trait association was evaluated with DNA marker and SNP/INDEL analyses using General Linear Model (GLM) and LS means/pdiff mean separation procedures using statistical package SAS 9.4 (SAS institute, NC, Cary, USA). DNA markers showing polymorphic bands were subjected to alle-

Table 2: DNA markers used in the present study.

Marker	Location related to <i>Ghd7</i>	Forward primer (5'→3')	Reverse primer (5'→3')	Ta*°C	Reference
<i>Seq1-2</i>	Within <i>Ghd7</i>	GCAAGGGGATGTCTAAACGA	AATTTTTGACCGTCGGATTC	53	Lu et al., (2012)
<i>Seq7-8</i>		CATACGGATCCAGCCTCTGT	TTGCAATGATGCGTATTAC	54	
<i>RM1135</i>	SSRs Flanking <i>Ghd7</i>	AGCCAACCAAGCAAGATAGC	ACACACATGTAAGCCTCCCC	58	Grameme QTL Database, (2017)
<i>RM5499</i>		TGGAGTACGACGTGATCGTG	CAGAAACGGGAGGGGATC	57	
<i>RM5346</i>		TGCCTCACGATGGTTCGAG	CTTCGTCCACCCAATTTGAC	56	
<i>RM5436</i>		CAAAGGGGGTGTCTCTATG	GTTGCTCGTCCTACATGTGC	57	
<i>G7rq</i>	Linked to <i>Ghd7</i>	AGGTGCTACGAGAAGCAAATCC	GGGCCTCATCTCGGCATAG	59	Xue et al., (2008)
<i>K20#</i>	Monomorphic locus	TCAGGTGATGGGAATCATTG	TGTTCCAACCAACAACCTG	55	Chin et al., (2010)

*Ta- Primer annealing temperature

- Monomorphic marker to detect the quality of DNA and positivity of PCR

scoring. A score of '1' was assigned for the presence of the band and a score of '0' was assigned for the absence of the band. An unrooted UPGMA tree was constructed based on the binary data of the polymorphic marker alleles using Nei-Gobi method in PHYLIP software package (version 3.6) (Felsenstein, 2005). The polymorphic band data were subjected to single marker analysis separately with the yield traits; overall yield, SN, seed weight, seed width, SL, EL, EW, HSW and HEW by employing GLM procedure in SAS. The association between the yield class and the marker alleles was analyzed using cross tabulation and chi-square test in Minitab16.0 (Minitab Inc., USA).

SNP/INDEL polymorphism at Seq7-8 and its association with yield traits

The DNA sequences of *Seq7-8* locus were aligned using ClustalW algorithm (Thompson *et al.*, 1997) built in MEGA V7 (Kumar *et al.*, 2016). The sequence alignment was used to identify the SNPs and INDELS present within the locus. The allelic states at polymorphic base positions were identified for cultivars. The alignment was used to draw the Neighbor-Joining tree using Tamura-Nei model of selection in MEGA V7 (Kumar *et al.*, 2016). The gaps were treated as complete deletions and both transversions and transitions were treated as substitutions. The resulting tree was visualized and edited using FigTree v1.4.3 (Rambaut, 2014). The allelic data obtained from the sequence polymorphism were subjected to single marker analysis. The associations between the SNP/INDEL allelic states at each variable position and the yield traits including overall yield, SN, seed weight, SL, seed width, EL, EW, HSW and HEW were analyzed by using GLM procedure in SAS.

RESULTS AND DISCUSSION

DNA length polymorphism of *Ghd7*

For the 12 rice cultivars, 13 alleles with an average of 1.86 alleles per locus were detected for seven DNA markers. The mean number of alleles detected in the present study is lower than results reported in similar previous studies Nagaraju *et al.*, (2002); Joshi and Behera, (2006) and Herrera *et al.*, (2008) (3.8, 4.58, 2.6 and 4.23 alleles respectively). However, the number of alleles detected per locus in the present study is comparable with the study reported by Singh *et al.*, (2004). Out of the seven DNA markers, four (*RM5346*, *RM5436*, *Seq1-2* and *Seq7-8*) were polymorphic. However, all the seven DNA markers gave successful amplifications for all 12 cultivars. Furthermore, in this study the number of alleles detected per locus was in the range of one to three. The minimum number of alleles (one) was detected in three loci; *RM 5499*, *RM 1135* and *G7rq*. Two alleles were detected at three loci; *RM5346*, *RM5436* and *Seq7-8* while three alleles (the highest polymorphism) were detected for *Seq1-2* marker.

Figure 1 shows the allelic diversity of the seven DNA markers within *Ghd7* locus for the 12 rice cultivars. The cultivar Bg 90-2 was used as the reference haplotype since it could be identified as the highest yielding genotype (data recorded at RRDI, Bathalagoda). Accordingly, ten marker haplotypes could be identified with the reference haplotype

of Bg 90-2. It was noticed that, the rice cultivars under the same yield category did not fall under the same haplotype. Instead, the high-yielding: Bg 366, moderately-yielding: Bg 310 and low-yielding: Bw 272-6b cultivars fell under the same haplotype, thus reflecting a close genetic relationship among these three cultivars regardless of their yield status. In contrast, other nine individual cultivars fell under nine distinct marker haplotypes, thereby revealing a higher genetic divergence among them.

Furthermore, the unrooted unweighted neighbor-joining tree constructed based on DNA band data, categorized the 12 rice cultivars as shown in Figure 1. Accordingly, Bg 366, At 307, Bg 310, Bw 367 and Bw 272-6b were grouped into one cluster. Furthermore, At 362 and Bg 300 claded together while Bg 352 and *Pachchaperumal* clustered together. The cultivars within the same cluster suggest their close genetic relationship. However, Bg 90-2, Bg 250 and *Suwadhal* remained un-clustered with three unique haplotypes, revealing their genetic divergence with respect to all the other cultivars (Figure 1). Moreover, the two landraces; *Pachchaperumal* and *Suwadhal* had a less genetic similarity between them which is also reflected by marker-haplotype analysis in which they fell under two distinct haplotypes. Furthermore, it was noticed that all the rice cultivars belonging to the same yield category was not co-clustered, deducing a low genomic similarity among them and revealing that *Ghd7* is not the only genetic player in deciding the rice yield.

Association between the polymorphic DNA markers and the yield related traits

The genetic polymorphism among the rice cultivars further gave an insight into marker-trait associations which facilitates the marker assisted selection (MAS), where superior cultivars are selected with the use of molecular markers. The results of marker-trait association in *Maha* season and *Yala* season are given in Table 3 and Table 4 respectively ($P < 0.05$). A total of 30 significant marker-trait associations could be detected in *Maha* season, while it was only 25 for the *Yala* season ($P < 0.05$). These results vary from those in similar previous studies reporting multiple seasons (Vanniarajan *et al.*, 2012). Such discrepancies might be due to the differences in number and the variations of selected rice cultivars, yield related traits, selected markers and the environment in which the particular study was conducted. Incompatibility of the results between the two seasons might be due to the environmental factors such as temperature, light conditions which affect the overall yield complicating the process of MAS (Schmierer *et al.*, 2004; Reyna and Sneller, 2001).

In *Maha* season, out of the four polymorphic markers, *Seq7-8* marker was detected to be significantly associated with seven traits; overall yield, SL, EL, seed weight, width, SN and EW. In *Yala* season, both the *RM5436* and *RM5346* markers were significantly associated with six traits; overall yield, SN, seed weight, seed width, EW, SL and EL, seed weight, SL, seed width, EL and EW respectively. Out of the polymorphic DNA markers, only *RM5346* and *Seq7-8* markers were shown to have significant associations with different yield traits in both *Maha* and *Yala* seasons. DNA

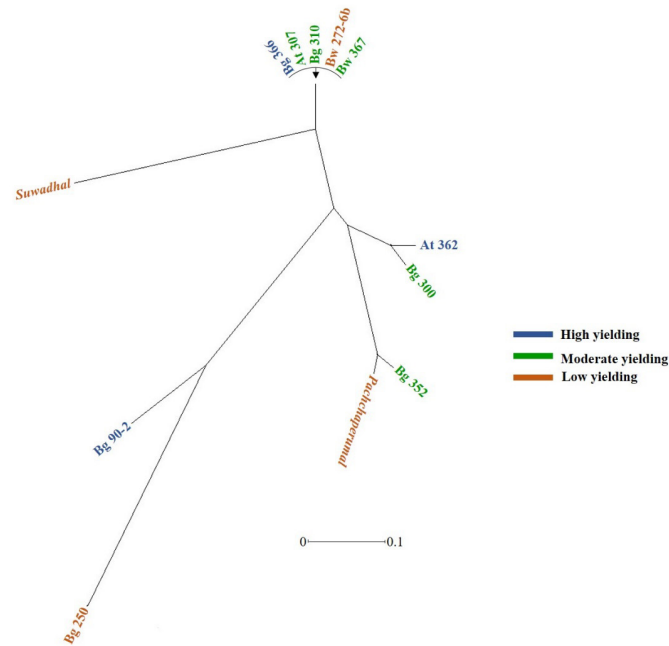


Figure 1: Unrooted unweighted neighbor-joining tree constructed for rice cultivars based on polymorphic marker loci linked to the *Ghd7* QTL region using Darwin software based on Jaccard coefficient (Perrier *et al.*, 2003). The yield status is indicated at the right side of the figure.

Table 3: Marker-trait associations for yield traits with DNA markers used in this study in *Maha* season.

Marker	Allele (bp)	Presence/absence of the allele	Trait								
			Yield (g)	Seed number	Seed weight (g)	Seed length (cm)	Seed width (cm)	Endosperm length (cm)	Endosperm width (cm)	100 seed weight (g)	100 endosperm weight (g)
RM5436	234	1	5.08 ^a	238.06 ^a	0.31 ^b	0.74 ^b	0.35 ^a	0.54 ^a	0.26 ^a	2.39 ^a	1.81 ^b
		0	5.09 ^a	201.81 ^a	0.40 ^a	0.88 ^a	0.30 ^b	0.64 ^a	0.25 ^a	2.71 ^a	2.22 ^a
RM5436	225	1	5.09 ^a	201.81 ^a	0.40 ^a	0.88 ^a	0.30 ^b	0.64 ^a	0.25 ^a	2.71 ^a	2.22 ^a
		0	5.08 ^a	238.06 ^a	0.31 ^b	0.74 ^b	0.35 ^a	0.54 ^a	0.26 ^a	2.39 ^a	1.81 ^b
RM5346	226	1	5.20 ^a	232.78 ^a	0.32 ^b	0.76 ^b	0.35 ^a	0.56 ^a	0.26 ^a	2.47 ^a	1.89 ^b
		0	3.77 ^b	258.36 ^a	0.28 ^b	0.62 ^b	0.26 ^b	0.42 ^b	0.25 ^a	1.50 ^b	1.44 ^b
RM5346	215	1	3.77 ^b	258.36 ^a	0.28 ^b	0.62 ^b	0.26 ^b	0.42 ^b	0.25 ^a	1.50 ^b	1.44 ^b
		0	5.20 ^a	232.78 ^a	0.32 ^b	0.76 ^b	0.35 ^a	0.56 ^a	0.26 ^a	2.47 ^a	1.89 ^a
Seq1-2	750	1	4.91 ^a	194.77 ^b	0.34 ^b	0.77 ^b	0.31 ^b	0.55 ^a	0.29 ^a	2.67 ^a	1.62 ^b
		0	5.11 ^a	243.75 ^a	0.31 ^b	0.75 ^b	0.35 ^a	0.55 ^a	0.26 ^a	2.36 ^a	1.89 ^a
Seq1-2	625	1	4.86 ^a	256.73 ^a	0.31 ^b	0.68 ^b	0.40 ^a	0.50 ^b	0.26 ^a	2.11 ^b	1.70 ^b
		0	5.30 ^a	214.76 ^b	0.32 ^b	0.83 ^a	0.29 ^b	0.59 ^a	0.27 ^a	2.66 ^a	2.00 ^a
Seq7-8	900	1	5.56 ^a	210.26 ^a	0.12 ^c	0.88 ^a	0.28 ^b	0.65 ^a	0.25 ^a	2.78 ^a	2.27 ^a
		0	4.98 ^a	240.05 ^a	0.09 ^c	0.73 ^b	0.35 ^a	0.53 ^a	0.26 ^a	2.33 ^a	1.76 ^b
Seq7-8	850	1	4.98 ^a	240.05 ^a	0.09 ^c	0.73 ^b	0.35 ^a	0.53 ^a	0.26 ^a	2.33 ^a	1.76 ^b
		0	5.56 ^a	210.26 ^a	0.12 ^c	0.88 ^a	0.28 ^b	0.65 ^a	0.25 ^a	2.78 ^a	2.27 ^a
GI	790	1	5.13 ^a	228.75 ^a	0.34 ^b	0.77 ^b	0.44 ^a	0.57 ^a	0.26 ^a	2.46 ^a	1.99 ^a
		0	5.04 ^a	239.22 ^a	0.30 ^b	0.74 ^b	0.29 ^b	0.53 ^a	0.27 ^a	2.38 ^a	1.75 ^b
GI	720	1	4.95 ^a	230.72 ^a	0.32 ^a	0.76 ^b	0.36 ^a	0.55 ^a	0.26 ^b	2.40 ^a	1.94 ^a
		0	5.46 ^a	247.50 ^a	0.29 ^b	0.74 ^b	0.29 ^b	0.56 ^a	0.27 ^a	2.47 ^a	1.58 ^b
GI	485	1	4.95 ^a	230.72 ^a	0.32 ^b	0.76 ^b	0.36 ^a	0.55 ^a	0.26 ^a	2.40 ^a	1.94 ^a
		0	5.46 ^a	247.50 ^a	0.29 ^b	0.74 ^b	0.29 ^b	0.56 ^a	0.27 ^a	2.47 ^a	1.58 ^b

Means denoted by the same letters within the column are not significantly different at $P < 0.05$. Presence of the allele is shown as 1 and absence is shown as 0.

Table 4: Marker-trait association for rice related traits with DNA markers used in this study in *Yala* season.

Marker	Allele (bp)	Presence/absence of the allele	Trait								
			Yield (g)	Seed number	Seed weight (g)	Seed length (cm)	Seed width (cm)	Endosperm length (cm)	Endosperm width (cm)	100 seed weight (g)	100 endosperm weight (g)
<i>RM5436</i>	234	1	5.36 ^a	236.78 ^a	0.18 ^a	0.76 ^b	0.27 ^a	0.54 ^b	0.25 ^a	2.42 ^a	1.76 ^a
		0	4.68 ^b	162.93 ^b	0.06 ^b	0.86 ^a	0.25 ^b	0.61 ^a	0.22 ^b	2.45 ^a	2.09 ^a
<i>RM5436</i>	225	1	4.68 ^b	162.93 ^b	0.06 ^b	0.86 ^a	0.25 ^b	0.61 ^a	0.22 ^b	2.45 ^a	2.09 ^a
		0	5.36 ^a	236.78 ^a	0.18 ^a	0.76 ^b	0.27 ^a	0.54 ^b	0.25 ^a	2.42 ^a	1.76 ^a
<i>RM5346</i>	226	1	5.47 ^a	230.87 ^a	0.18 ^a	0.78 ^a	0.27 ^a	0.56 ^a	0.25 ^a	2.49 ^a	1.85 ^a
		0	3.56 ^b	242.69 ^a	0.11 ^b	0.57 ^b	0.23 ^b	0.40 ^b	0.22 ^b	1.30 ^a	1.10 ^a
<i>RM5346</i>	215	1	3.56 ^b	242.69 ^a	0.11 ^b	0.57 ^b	0.23 ^b	0.40 ^b	0.22 ^b	1.30 ^a	1.10 ^a
		0	5.47 ^a	230.87 ^a	0.18 ^a	0.78 ^a	0.27 ^a	0.56 ^a	0.25 ^a	2.49 ^a	1.85 ^a
<i>Seq1-2</i>	750	1	5.61 ^a	222.99 ^a	0.19 ^a	0.77 ^a	0.29 ^a	0.56 ^a	0.28 ^a	2.64 ^a	1.90 ^a
		0	5.26 ^a	233.54 ^a	0.17 ^a	0.76 ^a	0.26 ^b	0.55 ^a	0.24 ^b	2.38 ^a	1.76 ^a
<i>Seq1-2</i>	625	1	5.04 ^a	249.47 ^a	0.14 ^b	0.69 ^b	0.27 ^a	0.49 ^b	0.24 ^a	2.09 ^b	1.51 ^b
		0	5.60 ^a	214.57 ^b	0.20 ^a	0.84 ^a	0.27 ^a	0.60 ^a	0.25 ^a	2.69 ^a	2.06 ^a
<i>Seq7-8</i>	900	1	5.09 ^a	184.34 ^b	0.18 ^a	0.90 ^a	0.26 ^b	0.64 ^a	0.23 ^b	2.79 ^a	2.26 ^a
		0	5.36 ^a	240.91 ^a	0.17 ^a	0.74 ^b	0.27 ^a	0.53 ^b	0.25 ^a	2.34 ^a	1.69 ^a
<i>Seq7-8</i>	850	1	5.36 ^a	240.91 ^a	0.17 ^a	0.74 ^b	0.27 ^a	0.53 ^b	0.25 ^a	2.34 ^a	1.69 ^a
		0	5.09 ^a	184.34 ^b	0.18 ^a	0.90 ^a	0.26 ^b	0.64 ^a	0.23 ^b	2.79 ^a	2.26 ^a
<i>GI</i>	790	1	5.09 ^a	209.71 ^b	0.16 ^b	0.77 ^a	0.26 ^b	0.56 ^a	0.24 ^b	2.38 ^a	1.82 ^a
		0	5.47 ^a	247.20 ^a	0.18 ^a	0.76 ^a	0.27 ^a	0.55 ^a	0.25 ^a	2.46 ^a	1.76 ^a
<i>GI</i>	720	1	5.11 ^b	222.38 ^a	0.18 ^a	0.76 ^a	0.27 ^a	0.55 ^a	0.24 ^a	2.36 ^a	1.75 ^a
		0	5.91 ^a	259.68 ^a	0.15 ^b	0.78 ^a	0.27 ^a	0.56 ^a	0.25 ^a	2.60 ^a	1.88 ^a
<i>GI</i>	485	1	5.11 ^b	222.38 ^a	0.18 ^a	0.76 ^a	0.27 ^a	0.55 ^a	0.24 ^a	2.36 ^a	1.75 ^a
		0	5.91 ^a	259.68 ^a	0.15 ^b	0.78 ^a	0.27 ^a	0.56 ^a	0.25 ^a	2.60 ^a	1.88 ^a

Means denoted by the same letters within the column are not significantly different at $P < 0.05$. Presence of the allele is shown as 1 and absence is shown as 0.

Table 5: Association between the yield class and the marker allele by Chi-square test analysis.

Marker	Allele (bp)	Pearson chi-square	P-value	Cramer's V-square
<i>RM5436</i>	234	2.18	- *	0.18
<i>RM5436</i>	225	2.18	-	0.18
<i>RM5346</i>	226	2.18	-	0.18
<i>RM5346</i>	215	2.18	-	0.18
<i>Seq1-2</i>	750	1.20	-	0.10
<i>Seq1-2</i>	625	2.00	0.37	0.17
<i>Seq7-8</i>	900	1.20	-	0.10
<i>Seq7-8</i>	850	1.20	-	0.10
<i>GI</i>	790	0.69	0.71	0.06
<i>GI</i>	720	2.67	0.26	0.22
<i>GI</i>	485	2.67	0.26	0.22

*Non-estimable: Only one individual in one genotypic class

marker *RM5346* was significantly associated with overall yield, seed weight, SL, seed width, EL, EW and *Seq7-8* was significantly associated with SL, EL, SN, seed width and EW in both the seasons ($P < 0.05$). This suggests that *RM5346* and *Seq7-8* markers could be more useful than other tested markers in implementing MAS for the studied rice cultivars.

The association between the polymorphic marker allele and the yield category (high, moderate and low), analyzed by cross tabulation and chi-square test analyses are given in the Table 5 ($P < 0.05$). The Cramer's V-square coefficient revealed that there was a weak association between the marker allele and the yield category (Cramer's V-square was closer to 0). This reveals that although the type of the DNA marker allele has considerable effects on the overall yield of the rice cultivars, it has no significant effect on the yield category ($P < 0.05$).

SNP/INDEL polymorphism at *Seq7-8* locus

The analyses of DNA length polymorphisms further facilitated the sequencing of rice cultivars at *Seq7-8* locus. This locus was specifically chosen due to its significant polymorphism. The nucleotide variations in the *Seq7-8* locus for 12 rice cultivars are shown in Figure 2. When all the 12 DNA sequences were aligned, a total of 23 nucleotide polymorphic positions including 11 single nucleotide polymorphisms (SNPs) were detected. At each SNP position, bi-allelic SNPs were only detected and accordingly, ten haplotypes could be detected. Moreover, it could be observed that Bg 90-2 and Bg 352 were falling under the same haplotype reflecting a close genetic relationship while At 307 and Bg 310 were falling under a separate haplotype reflecting their genetic divergence (Figure 3). The present findings revealed that the polymorphic SNPs and INDELs among these 12 rice cultivars in *Seq7-8* locus could be successfully employed in predicting the genetic diversity. The sequence alignment and identification of 23 nucleotide polymorphic positions at *Seq7-8* locus enabled the construction of an unrooted neighbor-joining tree (Figure 3). Accordingly, Bg 90-2 clustered with Bg 352 and At 307 clustered with Bg 310. This high genetic similarity between these two pairs of rice cultivars is also confirmed by sequence alignment (Figure 2).

Association between *Ghd7* sequence polymorphism and yield traits

The rice yield related traits evaluated in this study and the state of the association of these traits with SNP/INDELs in *Maha* and *Yala* seasons are depicted in Tables 6 and 7 respectively ($P < 0.05$). The association analysis revealed that only some of the SNPs/INDELs were significantly associated with rice yield traits (Table 6 and 7). Accordingly, in *Maha* season, only four SNPs among 23 nucleotide variations showed strong associations with the overall yield ($P < 0.05$). It was noticed that the overall yield was significantly high in the presence of G, A, T and G bases at 747, 862, 876 and 898 base positions respectively. Moreover, all SNPs and INDELs were strongly associated with seed width. These strong associations elucidate a high

contribution of nucleotide variation on the development of seeds ($P < 0.05$). In *Yala* season, a total of two SNPs and three INDELs were shown to be having significant associations with the overall yield ($P < 0.05$). A significantly higher overall yield was observed in SNPs at: 747th (G), 876th (T), and 889th (G), when the respective allele was present. Moreover, a higher yield was reported in *Yala* season for INDELs at: 870th (A) and 894th (A) positions, when the corresponding alleles were absent ($P < 0.05$). Furthermore, the highest number of significant associations (23 and 20) could be detected between seed width and polymorphic SNP/INDEL variations in both *Maha* and *Yala* seasons respectively ($P < 0.05$). However, the number and the type of the remaining associations were highly variable between the two seasons. Therefore, it was evident that due to the variations in yield parameters shown in two different seasons, the association results showed variations which could in turn validate environmental influence attributed to the variations in the overall yield (Table 6 and Table 7). Thus, these dissimilarities of the results between the two seasons might be complicating the process of MAS as observed in Moreau *et al.*, (2004a); Moreau *et al.*, (2004b); Schmierer *et al.*, (2004); Reyna and Sneller, (2001).

However, no significant associations could be detected between any of the nucleotide variations and the traits HSW and HEW in both seasons ($P < 0.05$). This fact explicates that the nucleotide variation of *Seq7-8* locus has no influence on HSW and HEW measurements. This result is in agreement with marker-trait association analysis in which, both the type of the *Seq7-8* marker allele and the presence/absence of the respective allele did not influence on HSW and HEW measurements. In consolation, Suresh *et al.*, (2017) has analyzed the associations between the whole sequence of *Ghd7* gene and grain yield parameters in a set of *indica* rice genotypes belonging to high and low yielding categories and results are quite comparable with our findings. Previously, the nucleotide diversity of the whole *Ghd7* gene and its associations with rice yield traits have been reported for different rice cultivars as mentioned above. However, in this study we analyzed the nucleotide diversity of *Ghd7* and its associations with yield traits for a set of Sri Lankan rice cultivars. Since none of the *Ghd7* related studies have been conducted for Sri Lankan rice germplasm, the results of the present study cannot be directly compared with those of the other studies. However, all these findings from different studies collectively give a deep insight into the features of *Ghd7* QTL and its contribution on yield performance of rice.

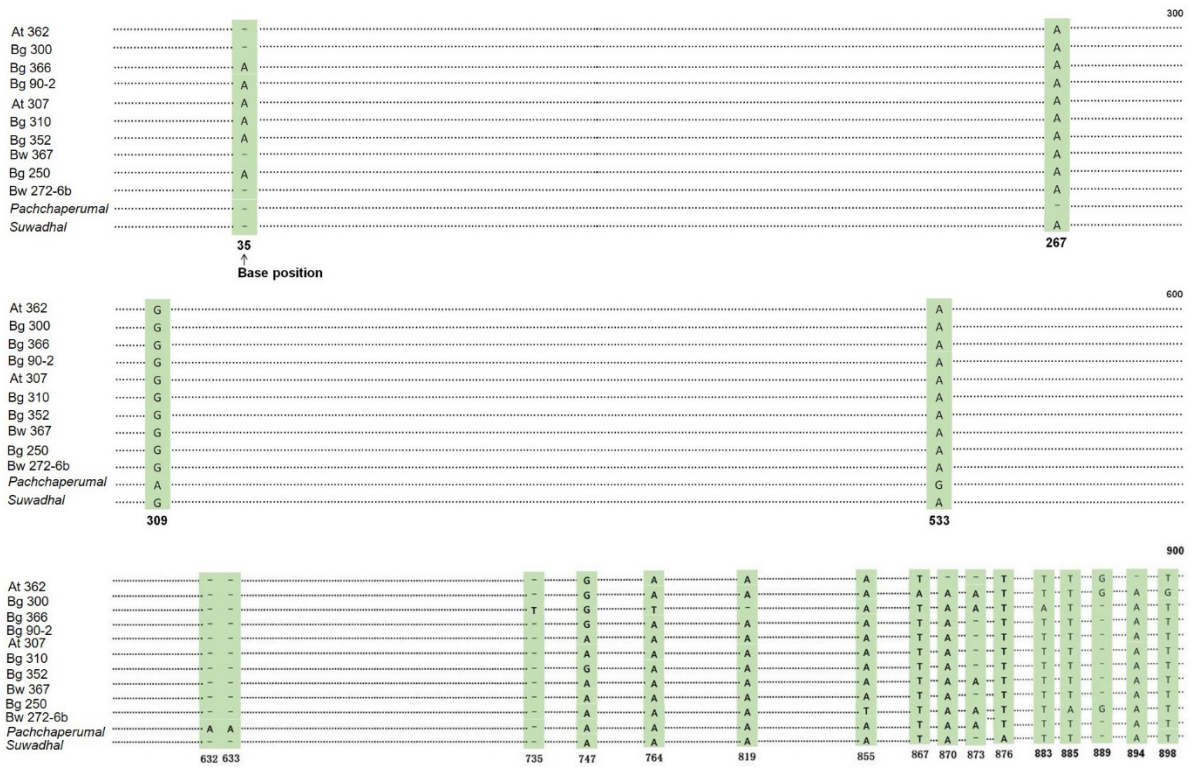


Figure 2: DNA sequence alignment for *Seq7-8* locus of the *Ghd7* QTL for 12 rice cultivars. The symbol ‘.’ Indicates monomorphic nucleotides and green color shades indicate SNP/INDEL positions across the cultivars. Ten haplotypes were identified based on 23 nucleotide variable positions.

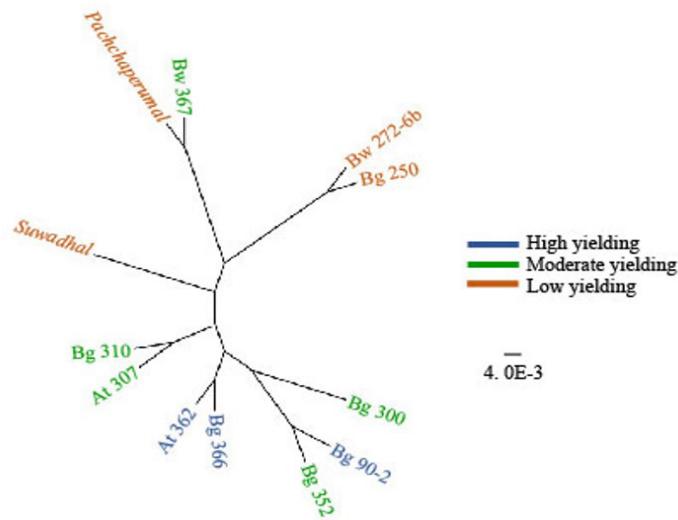


Figure 3: Unrooted neighbor-joining tree showing the genetic similarity among 12 rice cultivars based on the alignment of 12 sequences of *Seq7-8* locus using Tamura-Nei model of selection. The resulting tree was visualized and edited using FigTree v1.4.3(Rambaut, 2014).

Table 6: SNP/INDEL-yield trait associations for rice yield traits with sequence polymorphism at *Seq7-8* locus in *Maha* season.

Base position	SNP/INDEL	Allelic composition	Trait								
			Yield (g)	Seed number	Seed weight (g)	Seed length (cm)	Seed width (cm)	Endosperm length (cm)	Endosperm width (cm)	100 seed weight (g)	100 endosperm weight (g)
35	INDEL	A	5.20 ^a	212.46 ^b	0.33 ^a	0.80 ^a	0.42 ^a	0.59 ^a	0.27 ^a	2.59 ^a	1.96 ^a
		-	4.96 ^a	259.51 ^a	0.30 ^b	0.71 ^b	0.28 ^b	0.50 ^b	0.26 ^b	2.21 ^a	1.74 ^a
267	INDEL	A	5.09 ^a	239.28 ^a	0.31 ^b	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^b	2.39 ^a	1.83 ^a
		-	4.94 ^a	190.79 ^b	0.38 ^a	0.77 ^a	0.30 ^b	0.55 ^a	0.29 ^a	2.70 ^a	2.08 ^a
309	SNP	G	5.09 ^a	239.28 ^a	0.31 ^b	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^b	2.39 ^a	1.83 ^a
		A	4.94 ^a	190.79 ^b	0.38 ^a	0.77 ^a	0.30 ^b	0.55 ^a	0.29 ^a	2.70 ^a	2.08 ^a
533	SNP	A	5.09 ^a	239.28 ^a	0.31 ^b	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^b	2.39 ^a	1.83 ^a
		G	4.94 ^a	190.79 ^b	0.38 ^a	0.77 ^a	0.30 ^b	0.55 ^a	0.29 ^a	2.70 ^a	2.08 ^a
632	INDEL	A	4.94 ^a	190.79 ^b	0.38 ^a	0.77 ^a	0.30 ^b	0.55 ^a	0.29 ^a	2.70 ^a	2.08 ^a
		-	5.09 ^a	239.28 ^a	0.31 ^b	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^b	2.39 ^a	1.83 ^a
633	INDEL	A	4.94 ^a	190.79 ^b	0.38 ^a	0.77 ^a	0.30 ^b	0.55 ^a	0.29 ^a	2.70 ^a	2.08 ^a
		-	5.09 ^a	239.28 ^a	0.31 ^b	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^b	2.39 ^a	1.83 ^a
735	INDEL	T	5.05 ^a	224.81 ^a	0.36 ^a	0.75 ^a	0.27 ^b	0.57 ^a	0.26 ^b	2.36 ^a	1.92 ^a
		-	5.08 ^a	235.74 ^a	0.31 ^b	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^b	2.42 ^a	1.84 ^a
747	SNP	G	5.36 ^a	224.70 ^a	0.30 ^b	0.82 ^a	0.29 ^b	0.59 ^a	0.26 ^b	2.58 ^a	1.92 ^a
		A	4.88 ^b	242.30 ^a	0.33 ^a	0.71 ^b	0.39 ^a	0.52 ^b	0.26 ^b	2.28 ^a	1.80 ^a
764	SNP	T	5.05 ^a	224.81 ^a	0.36 ^a	0.75 ^a	0.27 ^b	0.57 ^a	0.26 ^b	2.36 ^a	1.92 ^a
		A	5.08 ^a	235.74 ^a	0.31 ^b	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^b	2.42 ^a	1.84 ^a
791	INDEL	A	5.08 ^a	235.74 ^a	0.31 ^b	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^b	2.42 ^a	1.84 ^a
		-	5.05 ^a	224.81 ^a	0.36 ^a	0.75 ^a	0.27 ^b	0.57 ^a	0.26 ^b	2.36 ^a	1.92 ^a
819	INDEL	G	4.96 ^a	239.90 ^a	0.35 ^a	0.74 ^b	0.28 ^b	0.56 ^a	0.25 ^b	2.34 ^a	1.85 ^a
		-	5.12 ^a	233.13 ^a	0.30 ^b	0.76 ^a	0.37 ^a	0.55 ^a	0.27 ^a	2.44 ^a	1.85 ^a
855	SNP	A	5.11 ^a	229.34 ^a	0.32 ^a	0.88 ^a	0.30 ^b	0.56 ^a	0.26 ^b	2.46 ^a	1.89 ^a
		T	4.74 ^b	304.34 ^a	0.29 ^a	0.74 ^b	0.35 ^a	0.43 ^b	0.24 ^b	1.78 ^b	1.41 ^a
867	SNP	A	5.83 ^a	252.21 ^a	0.19 ^b	0.81 ^a	0.30 ^b	0.58 ^a	0.26 ^b	2.48 ^a	2.00 ^a
		T	5.01 ^a	233.29 ^a	0.32 ^a	0.75 ^b	0.35 ^a	0.55 ^a	0.26 ^b	2.41 ^a	1.83 ^a
870	INDEL	A	5.09 ^a	235.07 ^a	0.31 ^a	0.74 ^b	0.35 ^a	0.55 ^a	0.26 ^b	2.40 ^a	1.81 ^a
		-	5.01 ^a	231.94 ^a	0.32 ^a	0.89 ^a	0.28 ^b	0.57 ^a	0.25 ^b	2.56 ^a	2.22 ^a
873	INDEL	A	5.21 ^a	258.12 ^a	0.31 ^a	0.70 ^b	0.28 ^b	0.52 ^b	0.26 ^b	2.26 ^a	1.74 ^a
		-	4.99 ^a	219.46 ^a	0.32 ^a	0.79 ^a	0.39 ^a	0.57 ^a	0.26 ^b	2.52 ^a	1.93 ^a
876	SNP	T	5.20 ^a	232.78 ^a	0.32 ^a	0.76 ^a	0.35 ^a	0.56 ^a	0.26 ^b	2.47 ^a	1.89 ^a
		A	3.77 ^b	258.35 ^a	0.28 ^b	0.62 ^b	0.26 ^b	0.42 ^b	0.25 ^b	1.50 ^b	1.44 ^a
883	SNP	T	5.05 ^a	224.81 ^a	0.36 ^a	0.75 ^a	0.27 ^b	0.57 ^a	0.26 ^b	2.36 ^a	1.92 ^a
		A	5.08 ^a	235.74 ^a	0.31 ^b	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^a	2.42 ^a	1.84 ^a
885	SNP	T	5.11 ^a	229.34 ^a	0.32 ^a	0.88 ^a	0.30 ^b	0.56 ^a	0.26 ^b	2.46 ^a	1.89 ^a
		A	4.74 ^a	304.34 ^a	0.29 ^b	0.74 ^b	0.35 ^a	0.43 ^b	0.24 ^b	1.78 ^b	1.41 ^a
889	INDEL	G	5.19 ^a	261.11 ^a	0.28 ^b	0.76 ^a	0.28 ^b	0.53 ^a	0.25 ^b	2.32 ^a	1.88 ^a
		-	5.04 ^a	226.64 ^a	0.33 ^a	0.75 ^a	0.36 ^a	0.55 ^a	0.27 ^a	2.45 ^a	1.84 ^a
894	INDEL	A	5.09 ^a	235.07 ^a	0.31 ^a	0.74 ^b	0.35 ^a	0.55 ^a	0.26 ^b	2.40 ^a	1.81 ^a
		-	5.01 ^a	231.94 ^a	0.32 ^a	0.89 ^a	0.28 ^b	0.57 ^a	0.25 ^b	2.56 ^a	2.22 ^a
898	SNP	G	5.83 ^a	252.21 ^a	0.19 ^b	0.81 ^a	0.30 ^b	0.58 ^a	0.26 ^b	2.48 ^a	2.00 ^a
		T	5.01 ^a	233.29 ^a	0.32 ^a	0.75 ^a	0.35 ^a	0.55 ^a	0.26 ^b	2.41 ^a	1.83 ^a
902	INDEL	A	5.19 ^a	247.75 ^a	0.32 ^a	0.73 ^b	0.28 ^b	0.54 ^a	0.26 ^b	2.35 ^a	1.82 ^a
		-	4.97 ^b	222.54 ^a	0.31 ^b	0.78 ^a	0.41 ^a	0.56 ^a	0.26 ^b	2.48 ^a	1.88 ^a
903	SNP	T	5.11 ^a	229.34 ^a	0.32 ^a	0.88 ^a	0.30 ^b	0.56 ^a	0.26 ^b	2.46 ^a	1.89 ^a
		A	4.74 ^b	304.34 ^a	0.29 ^b	0.74 ^b	0.35 ^a	0.43 ^b	0.24 ^b	1.78 ^b	1.41 ^a

Means denoted by the same letters within the column are not significantly different at $P<0.05$

Table 7: SNP/INDEL-yield trait associations for rice yield traits with sequence polymorphism at *Seq7-8* locus in *Yala* season

Base position	SNP/INDEL	Allelic composition	Trait								
			Yield (g)	Seed number	Seed weight (g)	Seed length (cm)	Seed width (cm)	Endosperm length (cm)	Endosperm width (cm)	100 seed weight (g)	100 endosperm weight (g)
35	INDEL	A	5.20 ^a	196.24 ^b	0.17 ^a	0.87 ^a	0.42 ^a	0.59 ^a	0.22 ^a	2.57 ^a	1.98 ^a
		-	5.43 ^a	268.10 ^a	0.18 ^a	0.81 ^b	0.41 ^b	0.51 ^b	0.23 ^a	2.09 ^a	1.60 ^a
267	INDEL	A	5.38 ^a	234.48 ^a	0.17 ^a	0.84 ^a	0.41 ^b	0.55 ^a	0.23 ^a	2.30 ^a	1.78 ^a
		-	4.59 ^a	201.70 ^a	0.23 ^a	0.84 ^a	0.45 ^a	0.55 ^a	0.12 ^b	2.62 ^a	1.85 ^a
309	SNP	G	5.38 ^a	234.48 ^a	0.17 ^a	0.84 ^a	0.41 ^b	0.55 ^a	0.23 ^a	2.30 ^a	1.78 ^a
		A	4.59 ^a	201.70 ^a	0.23 ^a	0.84 ^a	0.45 ^a	0.55 ^a	0.12 ^b	2.62 ^a	1.85 ^a
533	SNP	A	5.38 ^a	234.48 ^a	0.17 ^b	0.84 ^a	0.41 ^b	0.55 ^a	0.23 ^a	2.30 ^a	1.78 ^a
		G	4.59 ^b	201.70 ^a	0.23 ^a	0.84 ^a	0.45 ^a	0.55 ^a	0.12 ^a	2.62 ^a	1.85 ^a
632	INDEL	A	4.59 ^b	201.70 ^a	0.23 ^a	0.84 ^a	0.45 ^a	0.55 ^a	0.12 ^a	2.62 ^a	1.85 ^a
		-	5.38 ^a	234.48 ^a	0.17 ^a	0.84 ^a	0.41 ^b	0.55 ^a	0.23 ^a	2.30 ^a	1.78 ^a
633	INDEL	A	4.59 ^b	201.70 ^a	0.23 ^a	0.84 ^a	0.45 ^a	0.55 ^a	0.12 ^b	2.62 ^a	1.85 ^a
		-	5.38 ^a	234.48 ^a	0.17 ^a	0.84 ^a	0.41 ^b	0.55 ^a	0.23 ^a	2.30 ^a	1.78 ^a
735	INDEL	T	4.90 ^b	213.49 ^a	0.16 ^a	0.83 ^a	0.41 ^b	0.55 ^a	0.26 ^a	2.28 ^a	1.71 ^a
		-	5.36 ^a	233.63 ^a	0.17 ^a	0.84 ^a	0.42 ^a	0.55 ^a	0.22 ^a	2.33 ^a	1.79 ^a
747	SNP	G	5.77 ^a	224.47 ^b	0.20 ^a	0.88 ^a	0.41 ^b	0.60 ^a	0.23 ^a	2.65 ^a	2.03 ^a
		A	4.96 ^b	237.78 ^a	0.15 ^b	0.80 ^b	0.42 ^a	0.51 ^b	0.22 ^a	2.10 ^a	1.61 ^a
764	SNP	T	4.90 ^b	213.49 ^a	0.16 ^a	0.83 ^a	0.41 ^b	0.55 ^a	0.26 ^a	2.28 ^a	1.71 ^a
		A	5.36 ^a	233.63 ^a	0.17 ^b	0.84 ^a	0.42 ^a	0.55 ^a	0.22 ^a	2.33 ^a	1.79 ^a
791	INDEL	A	5.36 ^a	233.63 ^a	0.17 ^b	0.84 ^a	0.42 ^a	0.55 ^a	0.22 ^a	2.33 ^a	1.79 ^a
		-	4.90 ^b	213.49 ^a	0.16 ^a	0.83 ^a	0.41 ^b	0.55 ^a	0.26 ^a	2.28 ^a	1.71 ^a
819	INDEL	G	5.33 ^a	247.88 ^a	0.13 ^b	0.82 ^b	0.41 ^b	0.54 ^a	0.26 ^a	2.12 ^a	1.69 ^a
		-	5.31 ^a	226.74 ^a	0.18 ^a	0.84 ^a	0.42 ^a	0.55 ^a	0.22 ^a	2.40 ^a	1.82 ^a
855	SNP	A	5.23 ^a	220.73 ^a	0.18 ^a	0.85 ^a	0.42 ^a	0.56 ^a	0.23 ^a	2.39 ^a	1.83 ^a
		T	6.21 ^a	358.17 ^a	0.13 ^b	0.74 ^b	0.41 ^b	0.44 ^b	0.24 ^a	1.64 ^b	1.28 ^b
867	SNP	A	4.74 ^b	193.10 ^b	0.21 ^a	0.85 ^a	0.41 ^b	0.55 ^a	0.22 ^a	2.63 ^a	1.89 ^a
		T	5.38 ^a	235.83 ^a	0.17 ^a	0.84 ^a	0.42 ^a	0.55 ^a	0.23 ^a	2.30 ^a	1.78 ^a
870	INDEL	A	5.12 ^b	227.37 ^a	0.17 ^a	0.83 ^a	0.42 ^a	0.54 ^a	0.22 ^a	2.30 ^a	1.75 ^a
		-	7.31 ^a	278.64 ^a	0.24 ^a	0.93 ^a	0.40 ^b	0.65 ^a	0.25 ^b	2.63 ^a	2.17 ^a
873	INDEL	A	5.27 ^a	258.76 ^a	0.17 ^a	0.80 ^b	0.42 ^a	0.51 ^b	0.22 ^a	2.18 ^a	1.60 ^a
		-	5.36 ^a	212.04 ^a	0.18 ^a	0.86 ^a	0.42 ^a	0.58 ^a	0.23 ^a	2.44 ^a	1.92 ^a
876	SNP	T	5.47 ^a	230.87 ^a	0.18 ^a	0.85 ^a	0.42 ^a	0.56 ^a	0.22 ^a	2.42 ^a	1.85 ^a
		A	3.56 ^b	242.69 ^a	0.11 ^b	0.69 ^b	0.37 ^b	0.40 ^b	0.30 ^a	1.30 ^b	1.10 ^a
883	SNP	T	5.36 ^a	233.63 ^a	0.17 ^a	0.84 ^a	0.42 ^a	0.55 ^a	0.22 ^a	2.33 ^a	1.79 ^a
		A	4.90 ^b	213.49 ^a	0.16 ^a	0.83 ^a	0.41 ^b	0.55 ^a	0.26 ^a	2.28 ^a	1.71 ^a
885	SNP	T	5.23 ^a	220.73 ^a	0.18 ^a	0.85 ^a	0.42 ^a	0.56 ^a	0.23 ^a	2.39 ^a	1.83 ^a
		A	6.21 ^a	358.17 ^a	0.13 ^b	0.74 ^b	0.41 ^b	0.44 ^b	0.24 ^a	1.64 ^b	1.28 ^b
889	INDEL	G	6.09 ^a	272.40 ^a	0.20 ^a	0.85 ^a	0.41 ^b	0.55 ^a	0.24 ^a	2.30 ^a	1.78 ^a
		-	5.03 ^b	217.51 ^a	0.16 ^b	0.83 ^a	0.42 ^a	0.55 ^a	0.22 ^a	2.34 ^a	1.79 ^a
894	INDEL	A	5.12 ^b	227.37 ^a	0.17 ^a	0.83 ^a	0.42 ^a	0.54 ^a	0.22 ^a	2.30 ^a	1.75 ^a
		-	7.31 ^a	278.64 ^a	0.24 ^a	0.93 ^a	0.40 ^b	0.65 ^a	0.25 ^a	2.63 ^a	2.17 ^a
898	SNP	G	4.74 ^b	193.10 ^b	0.21 ^a	0.85 ^a	0.41 ^b	0.55 ^a	0.22 ^a	2.63 ^a	1.89 ^a
		T	5.38 ^a	235.83 ^a	0.17 ^b	0.84 ^a	0.42 ^a	0.55 ^a	0.23 ^a	2.30 ^a	1.78 ^a

902	INDEL	A	5.19 ^a	245.58 ^a	0.16 ^b	0.81 ^b	0.42 ^a	0.53 ^b	0.23 ^a	2.22 ^a	1.68 ^a
		-	5.44 ^a	218.55 ^a	0.19 ^a	0.86 ^a	0.42 ^a	0.57 ^a	0.23 ^a	2.44 ^a	1.89 ^a
903	SNP	T	5.23 ^a	220.73 ^a	0.18 ^a	0.85 ^a	0.42 ^a	0.56 ^a	0.23 ^a	2.39 ^a	1.83 ^a
		A	6.21 ^a	358.17 ^a	0.13 ^b	0.74 ^b	0.41 ^b	0.44 ^b	0.24 ^a	1.64 ^b	1.28 ^b

Means denoted by the same letters within the column are not significantly different at $P < 0.05$

CONCLUSIONS

The genetic divergence among cultivars was well supported by DNA length polymorphism and SNP/INDEL polymorphisms. Based on DNA length polymorphism six distinct marker haplotypes could be recognized. The SNP/INDELs present in the *Seq7-8* locus also reveal a high genetic diversity among the genotypes, which was validated by the sequence based haplotype alignment. The analysis of SNP/INDEL polymorphism enabled the identification of ten haplotypes. Accordingly, Bg 90-2 and Bg 352 fall under one haplotype and At 307 and Bg 310 fall under another distinct haplotype suggesting a close genetic relationship among them. Association of SNP/INDEL polymorphism with yield traits revealed a higher contribution of sequence polymorphism on determination of seed width, but a lower contribution on HSW and HEW parameters. The strong associations between four polymorphic DNA markers (*RM5436*, *RM5346*, *Seq7-8* and *Seq1-2*) with yield traits (SN, seed weight, SL, seed width, EL, EW, HSW, HEW and overall yield) reveal their subsequent applicability in MAS.

ACKNOWLEDGEMENTS

Authors wish to thank University of Peradeniya Research Grant (URG/2016/59/S) and the Director, Rice Research and Development Institute (RRDI), Bathalagoda, Sri Lanka.

REFERENCES

- Bentota, A.P. (2009). New Crop Varieties Recommended by the Varietal Release Committee of the Department of Agriculture Sri Lanka, *Department of Agriculture Press, Gannoruwa, Sri Lanka*, 1-4.
- Chin, J.H., Lu, X., Stephan, M., Gamuyao, H.R., Ismail, A., Wissuwa, M. and Heuer, S. (2010). Development and application of gene-based markers for the major rice QTL *Phosphorus uptake 1*. *Theoretical and Applied Genetics* **120**(6):1073-1086.
- Department of Agriculture (DOA), Sri Lanka. (2015). Food security: Current food situation. Available from: <http://www.ips.lk/> (25 July 2017).
- Department of Agriculture (DOA), Sri Lanka. (2017). Foot prints in rice variety improvements and its impact on rice production in Sri Lanka. Available from: <http://www.goviya.lk/> (10 February 2017).
- Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Seattle, WA: Department of Genome Sciences, University of Washington.
- Hemachandra, K.H.D.J.K. and Amarasinghe, L.D. (2016). Screening of selected rice varieties for *Meloidogyne graminicola* infestation and potential of *Trichoderma viride* to control its infestation. *Proceedings of the International Research Symposium on Pure and Applied Sciences (IRSPAS), Faculty of Science, University of Kelaniya, Sri Lanka*, Pp.06.
- Herrera, T.G., Duque, D.P., Almeida, I.P., Nunez, G.T., Pieters, A.J., Martinez, C.P. and Tohme, J.M. (2008). Assessment of genetic diversity in Venezuelan rice cultivars using simple sequence repeats markers. *Electronic Journal of Biotechnology* **11**(5):3-4.
- Huang, X., Qian, Q., Liu, Z., Sun, H., He, S., Luo, D., Xia, G., Chu, C., Li, J. and Fu, X. (2009). Natural variation at the *DEP1* locus enhances grain yield in rice. *Nature Genetics* **41**(4):494-497.
- International Union for Conservation of Nature (IUCN), Sri Lanka (2016). Rice farming: back to some traditional practices. Available from: www.iucn.org/ (10 February 2017).
- Joshi, R.K. and Behera, L. (2006). Identification and differentiation of indigenous non-basmati aromatic rice genotypes of India using microsatellite markers. *African Journal of Biotechnology* **6**(4):348-354.
- Kumar, S., Strecher, G. and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology of Evolution* **33**(7):1870-1878.
- Liu, T., Liu, H., Zhang, H. and Xing, Y. (2013). Validation and characterization of *Ghd7. 1*, a major quantitative trait locus with pleiotropic effects on spikelets per panicle, Plant Height, and Heading Date in Rice (*Oryza sativa* L.). *Journal of Integrative Plant Biology* **55**(10):917-927.
- Lu, L., Yan, W., Xue, W., Shao, D. and Xing, Y. (2012). Evolution and association analysis of *Ghd7* in rice. *PLoS ONE* **7**(5); e34021.
- Matsubara, K., Yamanouchi, U., Nonoue, Y., Sugimoto, K., Wang, Z., Minobe, Y. and Yano, M. (2011). *Ghd3*, encoding a plant homeodomain finger-containing protein, is a critical promoter of rice flowering. *The Plant Journal* **66**(4):603-612.
- Moreau, L., Charcosset, A. and Gallais, A. (2004a). Experimental evaluation of several cycles of marker-assisted selection in maize. *Euphytica* **137**(1):111-118.
- Moreau, L., Charcosset, A. and Gallais, A. (2004b). Use of trial clustering to study QTL × environment effects for grain yield and related traits in maize. *Theoretical and Applied Genetics* **110**(1):92-105.
- Nagaraju, J., Kathirvel, M., Kumar, R.R., Siddiq, E.A. and Hasnain, S.E. (2002). Genetic analysis of traditional and evolved basmati and non-basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. *Proceedings of the National Academy of Sciences* **99**(9): 5836-5841.
- Nageswara-Rao, M., Kwit, C., Agarwal, S., Patton, M.T.,

- Skeen, J.A., Yuan, J.S., Manshardt, R.M. and Stewart-Jr, C.N. (2013). Sensitivity of a real-time PCR method for the detection of transgenes in a mixture of transgenic and non-transgenic seeds of papaya (*Carica papaya* L). *BMC Biotechnology* **13**(1): 69.
- Perrier, X., Flori, A. and Bonnot, F. (2003). Data analysis methods. In: P. Hamon, M. Seguin, X. Perrier and J.C. Glaszmann(Eds), Genetic diversity of cultivated tropical plants. Enfield, Science Publishers, Montpellier. 43-76.
- Priyantha, G.D.A., Weerasinghe, B.G.D.S. and Rasanjali, K.A.C. (2013). Iron toxicity tolerance of improved rice varieties in the low country wet zone of Sri Lanka. *Proceedings of the International Symposium on Agriculture and Environment, University of Ruhuna, Sri Lanka*, Pp. 428-432.
- Rambaut, A. (2014). FigTree, a graphical viewer of phylogenetic trees. Available from: <http://tree.bio.ed.ac.uk/software/figtree/> (25 July 2018).
- Reyna, N. and Sneller, C.H. (2001). Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. *Crop Science* **41**(4): 1317-1321.
- Rice Research and Development Institute (RRDI), Sri Lanka (2017). Rice cultivation. Available from: <http://www.doa.gov.lk/> (10 February 2017).
- Schmierer, D.A., Kandemir, N., Kudrna, D.A., Jones, B.L., Ullrich, S.E. and Kleinhofs, A. (2004). Molecular marker-assisted selection for enhanced yield in malting barley. *Molecular Breeding* **14**(4):463-473.
- Singh, R.K., Sharma, R.K., Singh, A.K., Singh, V.P., Singh, N.K., Tiwari, S.P. and Mohapatra, T. (2004). Stability of mapped sequence tagged microsatellite site markers for establishing distinctness, uniformity and stability in aromatic rice. *Euphytica* **135**:135-143.
- Suresh, E.N., Vamireddy, L.N., Reddy, N.P.E. and Sekhar, M.R. (2017). Gene expression and sequence analysis of known yield genes in high yielding varieties of rice (*Oryza sativa* L.). *Canadian Journal of Biotechnology* **1**:202-203.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The clustal x windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**(24):4876-4882.
- Vanniarajan, C., Vinod, K.K. and Pereira, A. (2012). Molecular evaluation of genetic diversity and association studies in rice (*Oryza sativa* L.). *Journal of Genetics* **91**(1):9-19.
- Weng, X., Wang, L., Wang, J., Hu, Y., Du, H., Xu, C., Xing, Y., Li, X., Xiao, J. and Zhang, Q. (2014). Grain number, plant height, and heading date 7 is a central regulator of growth, development, and stress response. *Plant Physiology* **164**: 735-747.
- Xing, Y., Zhang, Q. and Xue, W. (2014). Cloning and application of a pleiotropic gene *Ghd7* that controls grains yield, heading date and plant height of rice. US Patent No.: US8653329B2, Washington, DC: US Patent and Trademark Office.
- Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li, X. and Zhang, Q. (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nature Genetics* **40**(6):761-767.
- Yan, W.H., Wang, P., Chen, H.X., Zhou, H.J., Li, Q.P., Wang, C.R., Ding, Z.H., Zhang, Y.S., Yu, S.B., Xing, Y.Z. and Zhang, Q.F. (2011). A major QTL, *Ghd8*, plays pleiotropic roles in regulatory grain productivity, plant height, and heading date in rice. *Molecular Plant* **4**(2):319-330.
- Yang, L., Liu, T., Li, B., Sui, Y., Chen, J., Wing, R.A. and Chen, M. (2012). Comparative sequence analysis of the *Ghd7* in the grass genomes. *PLoS ONE* **7**(11): e50236.
- Yano, M. and Sasaki, T. (1997). Genetic and molecular dissection of quantitative traits in rice. *Plant Molecular Biology* **35**(1-2):145-153.
- Zhang, J., Zhou, X., Yan, W., Zhang, Z., Lu, L., Han, Z., Zhao, H., Liu, H., Song, P., Hu, Y., Shen, G., He, Q., Guo, S., Gao, G., Wang, G. and Xing, Y. (2015). Combinations of the *Ghd7*, *Ghd8* and *Hdl* genes largely define the ecogeographical adaptation and yield potential of cultivated rice. *New Phytologist* **208**:1056-1066.